Document Cover Sheet

Project Number				
Document Title	Skype Audio Specification v4.0.5			
Source	MWM Acoustics			
Contact	Name:	Glenn Hess	Phone:	317-596-1721
	Complete Address:	Suite 520 6602 East 75th Street Indianapolis, IN 46250	Fax:	317-849-8178
			Email:	hess@mwmacoustics.com
Distribution	TR-41.3.3			
Intended Purpose of Document (Select one)		For Incorporation Into TIA Publication		
	X	For Information		
		Other (describe) -		

The document to which this cover statement is attached is submitted to a Formulating Group or sub-element thereof of the Telecommunications Industry Association (TIA) in accordance with the provisions of Sections 6.4.1–6.4.6 inclusive of the TIA Engineering Manual dated March 2005, all of which provisions are hereby incorporated by reference.

Abstract

The attached Skype™ specification is drawing world-wide attention by audio product manufactures. This public domain document covers VoIP transmission test methods and performance requirements based exclusively on the Skype™ soft client. The requirements are divided into several groups covering handsets, headsets, speakerphones, and other audio devices such as cordless, DECT, and Bluetooth products. Telecom audio products must meet these audio requirements to be Skype™ certified. This specification could supersede TIA 810B and 920 for some product companies here in North America.

The Skype™ specification has three priority levels of audio performance identified as P1, P2, and P3, where P1 is a mandatory must comply requirement, P2 a should pass, and P3 nice or desirable to meet. The test conditions and/or requirement limits differ between the three priorities. Test parameters include send and receive frequency response, overall sensitivity, volume level, distortion, speech-to-noise, stability, crosstalk, echo, and ring tone loudness for normal band, wideband, and super wideband devices. These measurements are performed on an ITU-T compatible HATS with the Type 3.3 ear simulator.

Hardware Certification Audio Specification

Copyright © 2009 Skype. All Rights Reserved.

Last saved: 2009-04-01	Author: Markus Vaalgamaa Ergo Esken	Approved by: Ed Botterill
Status: **Final**		Version: 4.0.5
Filename: Test_SpecAudio_4.0.5.doc		
Security Classification: Public		

SUMMARY OF REVISIONS

Version	Date	Comments	Valid
4.0.5	**2009-04-01**	**Fixed some cross-references.**	**2009-04-01**
4.0.4	**2009-03-31**	**Added sub categories** General audio requirements - All groups: Additional requirements for PC or Mac accessories Headset audio UI: Audio performance requirements for Skype Super Wideband Certification Definitions and references moved to end of document.	**2009-04-01**
4.0.1	**2008-11-06**	**Few typos corrected, more explanations added based on comments by HeadAcoustics**	**2009-04-01**
4.0	**2008-10-01**	**Specification changes frozen. Changes are listed down in Appendix**	**2009-04-01**
3.0	**2008-01-01**	**Specification changes frozen.**	**2008-07-01**
2.2	**2007-12-31**	**List of major modifications:** **Modified requirement:** • Divided Additional delay to speech signal to receiving and sending direction requirements • Priority: 1 Minimum crosstalk from receiving to sending direction to Headset, Handset and Other Audio product groups **Added requirements:** • Error! Reference source not found. Error! Reference source not found. Error! Reference source not found. Error! Reference source not found. To headset, handset and speakerphone audio UI groups: • Priority: 1 Microphone - Sensitivity at loud speech level • Priority: 1 Microphone – Speech to self noise ratio during speech activity To speakerphone UI group: • Priority: 1,2 & 3 Microphone – Speech to background noise ratio	**2008-07-01**

Copyright © 2009 Skype Inc. All Rights Reserved.

CONTENTS

1. INTRODUCTION 6

1.1 PURPOSE 6
1.2 AUDIO UI GROUPS 6
1.2.1 Headset audio UI group 7
1.2.2 Handset audio UI group 7
1.2.3 Speakerphone audio UI group 7
1.2.4 Other audio product group 8
1.2.5 Non-audio product group 8
1.3 AUDIO REQUIREMENTS AND PRIORITIES – OVERVIEW 9
1.3.1 Audio performance 9
1.3.2 Quality expectation of the audio UI groups 9
1.3.3 Use of the test case priorities 9

2. GENERAL AUDIO REQUIREMENTS VALID FOR ALL GROUPS 10

2.1 ALL GROUPS: AUDIO PERFORMANCE REQUIREMENTS 10
2.1.1 Priority: 1 Round trip delay of speech signals 10
2.1.2 Priority: 1 Total quality loss in sending direction 10
2.1.3 Priority: 1 Total quality loss in receiving direction 11
2.2 ALL GROUPS: ADDITIONAL REQUIREMENTS FOR PC OR MAC ACCESSORIES 11
2.2.1 Priority: 1 Analog gain adjustment latency 11
2.2.2 Priority: 1 Device – Sampling frequency accuracy 12
2.3 GENERAL AUDIO TEST INSTRUCTIONS 12
2.3.1 Objective testing measurement setup 12

3. HEADSET AUDIO UI GROUP 14

3.1 HEADSET: AUDIO PERFORMANCE REQUIREMENTS 14
3.1.1 Priority: 1 Microphone – Sensitivity at normal speech level 14
3.1.2 Priority: 2 Microphone – Sensitivity at lowered speech level 14
3.1.3 Priority: 1 Microphone – Sensitivity at loud speech level 14
3.1.4 Priority: 1 Microphone – Frequency response 14
3.1.5 Priority: 2 Microphone – Frequency response 15
3.1.6 Priority: 1 Microphone – Speech to self noise ratio 16
3.1.7 Priority: 2 Microphone – Speech to self noise ratio 16
3.1.8 Priority: 3 Microphone – Speech to self noise ratio 17
3.1.9 Priority: 2 Microphone – Speech to self noise ratio during speech activity 17
3.1.10 Priority: 2 Microphone – Speech to background noise ratio 17
3.1.11 Priority: 1 Earpiece – Speech to self noise ratio 17
3.1.12 Priority: 2 Earpiece – Speech to self noise ratio 17
3.1.13 Priority: 3 Earpiece – Speech to self noise ratio 18
3.1.14 Priority: 1 Earpiece – Frequency response 18
3.1.15 Priority: 2 Earpiece – Frequency response 19
3.1.16 Priority: 1 Earpiece – Stability of frequency response 19
3.1.17 Priority: 2 Earpiece – Stability of frequency response 20
3.1.18 Priority: 3 Earpiece – Stability of frequency response 20
3.1.19 Priority: 1 Minimum crosstalk from receiving to sending direction 20
3.2 HEADSET: REQUIREMENTS FOR SKYPE SUPER WIDEBAND CERTIFICATION (OPTIONAL) 20
3.2.1 Priority: 1 Microphone – Frequency response 20
3.2.2 Priority: 1 Earpiece – Frequency response 21
3.2.3 Priority: 1 Earpiece – Speech to noise ratio 22
3.3 HEADSET: SUPPORTING AUDIO DOCUMENTATION REQUIREMENTS 22
3.3.1 Priority: 1 Verifying supporting documentation for Headset Audio UI group 23
3.4 HEADSET: AUDIO TEST INSTRUCTIONS 23
3.4.1 Objective testing measurement setup 23

Copyright © 2009 Skype Inc. All Rights Reserved.

4. HANDSET AUDIO UI GROUP 25

4.1 HANDSET: AUDIO PERFORMANCE REQUIREMENTS 25
4.1.1 Priority: 1 Microphone – Sensitivity at normal speech level 25
4.1.2 Priority: 2 Microphone – Sensitivity at lowered speech level 25
4.1.3 Priority: 1 Microphone – Sensitivity at loud speech level 25
4.1.4 Priority: 1 Microphone – Frequency response 25
4.1.5 Priority: 2 Microphone – Frequency response 26
4.1.6 Priority: 1 Microphone – Speech to self noise ratio 27
4.1.7 Priority: 2 Microphone – Speech to self noise ratio 27
4.1.8 Priority: 3 Microphone – Speech to self noise ratio 27
4.1.9 Priority: 2 Microphone – Speech to self noise ratio during speech activity 28
4.1.10 Priority: 2 Microphone – Speech to background noise ratio 28
4.1.11 Priority: 1 Earpiece – Speech to self noise ratio 28
4.1.12 Priority: 2 Earpiece – Speech to self noise ratio 28
4.1.13 Priority: 3 Earpiece – Speech to self noise ratio 28
4.1.14 Priority: 1 Earpiece – Frequency response 29
4.1.15 Priority: 2 Earpiece – Frequency response 29
4.1.16 Priority: 3 Earpiece – Frequency response 30
4.1.17 Priority: 1 Minimum crosstalk from receiving to sending direction 31
4.1.18 Priority: 1 Earpiece – Stability of frequency response 31
4.1.19 Priority: 2 Earpiece – Stability of frequency response 32
4.1.20 Priority: 3 Earpiece – Stability of frequency response 32
4.1.21 Priority: 1 Earpiece – Suitable volume level for office and home handset (Indoor) 32
4.1.22 Priority: 2 Earpiece – Suitable volume level for office and home handset (Indoor) 32
4.1.23 Priority: 1 Earpiece – Suitable volume level for “anywhere” handset (Outdoor) 32
4.1.24 Priority: 2 Earpiece – Suitable volume level for “anywhere” handset (Outdoor) 33
4.1.25 Priority: 1 Maximum ring tone loudness 33
4.1.26 Priority: 2 Maximum ring tone loudness 33
4.1.27 Priority: 3 Maximum ring tone loudness 34
4.2 HANDSET: SUPPORTING AUDIO DOCUMENTATION REQUIREMENTS 34
4.2.1 Priority: 1 Verifying supporting documentation for Handset audio 34
4.3 HANDSET: AUDIO TEST INSTRUCTIONS 35
4.3.1 Objective testing measurement setup 35

5. SPEAKERPHONE AUDIO UI GROUP 37

5.1 SPEAKERPHONE: AUDIO PERFORMANCE REQUIREMENTS 37
5.1.1 Priority: 1 Microphone – Sensitivity at normal speech level 37
5.1.2 Priority: 1 Microphone – Sensitivity at lowered speech level 37
5.1.3 Priority: 1 Microphone – Sensitivity at loud speech level 37
5.1.4 Priority: 1 Microphone – Frequency response 37
5.1.5 Priority: 2 Microphone – Frequency response 38
5.1.6 Priority: 3 Microphone – Frequency response 39
5.1.7 Priority: 1 Microphone – Speech to self noise ratio 40
5.1.8 Priority: 2 Microphone – Speech to self noise ratio 40
5.1.9 Priority: 3 Microphone – Speech to self noise ratio 41
5.1.10 Priority: 2 Microphone – Speech to self noise ratio during speech activity 41
5.1.11 Priority: 1 Amount of acoustic echo 41
5.1.12 Priority: 2 Amount of acoustic echo 41
5.1.13 Priority: 3 Amount of acoustic echo 42
5.1.14 Priority: 2 Echo loss in single talk during Skype call 42
5.1.15 Priority: 3 Echo loss in single talk without Skype speech improvements 43
5.1.16 Priority: 1 Loudspeaker – Frequency response 43
5.1.17 Priority: 2 Loudspeaker – Frequency response 44
5.1.18 Priority: 3 Loudspeaker – Frequency response 44
5.1.19 Priority: 1 Loudspeaker – Suitable volume level for quiet office use 45
5.1.20 Priority: 1 Loudspeaker – Distortion at quiet office use 45
5.1.21 Priority: 2 Loudspeaker – Suitable volume level for normal office use 46
5.1.22 Priority: 2 Loudspeaker – Distortion at normal office use 46
5.1.23 Priority: 3 Loudspeaker – Suitable volume level for noisy office use 46
5.1.24 Priority: 3 Loudspeaker – Distortion at noisy office use 46

Copyright © 2009 Skype Inc. All Rights Reserved.

5.1.25 Priority: 2 Loudspeaker – Volume level at maximum operating distance 47
5.1.26 Priority: 2 Microphone – Sensitivity at maximum operating distance 47
5.1.27 Priority: 3 Microphone – Speech to self noise ratio at maximum operating distance 47
5.2 SPEAKERPHONE: SUPPORTING AUDIO DOCUMENTATION REQUIREMENTS 47
5.2.1 Priority: 1 Verifying supporting documentation for Speakerphone audio 47
5.3 SPEAKERPHONE: AUDIO TEST INSTRUCTIONS 48
5.3.1 Objective testing measurement setup 48
5.3.2 Subjective testing measurement setup 49

6. OTHER AUDIO PRODUCT GROUP 51

6.1 OTHER AUDIO PRODUCT: AUDIO PERFORMANCE REQUIREMENTS 51
6.1.1 Priority: 1 Frequency responses – sending and receiving directions 51
6.1.2 Priority: 1 Product provides suitable levels for audio signal output 52
6.1.3 Priority: 1 Product provides suitable levels for audio signal input 52
6.1.4 Priority: 1 Minimum crosstalk from receiving to sending direction 52
6.2 OTHER AUDIO PRODUCT: SUPPORTING AUDIO DOCUMENTATION REQUIREMENTS 52
6.2.1 Priority: 1 Verifying supporting documentation for Other audio product 52
6.3 OTHER AUDIO PRODUCT: AUDIO TEST INSTRUCTIONS 53
6.3.1 Objective testing measurement setup 53

7. NON-AUDIO PRODUCT GROUP 54

7.1 NON-AUDIO PRODUCT: AUDIO PERFORMANCE REQUIREMENTS 54
7.1.1 Priority: 1 Continuous transmission of speech 54
7.1.2 Priority: 2 Continuous transmission of speech 54
7.2 NON-AUDIO PRODUCT: SUPPORTING AUDIO DOCUMENTATION 54
7.2.1 Priority: 1 Verifying supporting documentation for Non-audio product 54
7.3 NON AUDIO PRODUCT: AUDIO TEST INSTRUCTIONS 55
7.3.1 Objective testing measurement setup 55

8. LIST OF ENVIRONMENTS 56

8.1 LIST OF TEST PLATFORMS 56
8.1.1 Skype Audio Test Lab 56
8.1.2 Compatible testing environment 58

9. APPENDIX 59

9.1 DEFINITIONS 59
9.2 REFERENCES 64
9.3 CHANGES BETWEEN 4.0 AND 3.0 VERSIONS 64
9.3.1 Major changes 64
9.3.2 Introduction, Abbreviations and References 65
9.3.3 General audio requirements 65
9.3.4 Headset audio UI 66
9.3.5 Handset audio UI 66
9.3.6 Speakerphone audio UI 67
9.3.7 Other audio product 67
9.3.8 Non-audio product 68
9.3.9 List of environments 68

Copyright © 2009 Skype Inc. All Rights Reserved.

1. Introduction

This specification defines the audio requirements for Skype Certified Solutions. The requirements are divided into several groups, based on the acoustic user interface (UI) type.

For each group there are certain audio requirements. The requirements are mostly the same for all products that fall into one of the categories, but there can be small variances within one group, depending on the underlying technology.

In addition to the audio requirements, any product under test must comply with general Skype Certification Specifications which can be downloaded from Skype Developer Zone (https://developer.skype.com/Certification/Hardware/Specs/). A rule to calculate the final test result for a product is defined in Skype Certification Specifications.

1.1 Purpose

The requirements found in this test specification define the main parts of audio performance, ergonomic topics and documentation.

The purpose of this document is not to define requirements for all aspects of audio, but rather to concentrate on parts that affect the end user experience. Thus the tests cases based on these audio requirements do not replace other necessary testing that a vendor should and must perform in order to improve the end quality of the product before applying for Skype Certified label.

1.2 Audio UI groups

Skype Certified products are broken into several categories that are based on the acoustic interface type of the product. The groups are:

- Headset audio UI,
- Handset audio UI,
- Speakerphone audio UI,
- Other audio products
- No Acoustic UI audio product group.

One product can belong to several audio UI groups depending on possible usage scenarios of the product. For example: Wi-Fi phone, can have Handset, Headset and Speakerphone audio UI functionalities built into it, because it can have a handsfree feature (headset included in the package) and speakerphone mode support. In these cases requirements and test cases for several audio UI groups are valid.

Important point to notice is that some audio groups give actual acoustic interface to the user and others don't.

The groups that provide acoustic user interface are:

- Headset audio UI,
- Handset audio UI and
- Speakerphone audio UI groups.

Products belonging to these groups must have microphone or similar speech pickup device or loudspeaker / earpiece to reproduce speech, or even both.

Non-acoustic user interface groups are

Copyright © 2009 Skype Inc. All Rights Reserved.

- Other audio products
- Non-audio product group,

They include products that do not have microphone, earpiece or loudspeaker that would be used for communication. Examples: soundcard, ATA, motherboard.

1.2.1 Headset audio UI group

Headset audio UI product consists of two main components – earpiece(s) and microphone assembled together so that the headset can be fixed on the user's head or ear(s). Products that have microphone and earpieces separated physically (for example desktop microphone and headphones) also fall into Headset audio UI group.

Skype certification specifications for Headset audio UI group are categorized as follows:

Plug-in Headsets – wired headsets. They usually have standard 3.5 mm mini-plug audio connectors or USB cable.

Cordless Headsets – wireless headsets. They operate through radio frequencies, for example Bluetooth, DECT or Infrared.

Headset is connected to another device, like PC or PDA that has Skype running in it. Examples of Headset audio UI devices are illustrated below:

1.2.2 Handset audio UI group

A handset audio UI product is a handset that the user holds in his hand and puts next to his ear when in a call, so the form factor of the device is similar to that of a landline or mobile phone. The handset has both earpiece and microphone in the same device.

Just like the headset, handset can be wired or wireless. Skype certification specifications valid for this category are Plug-in Handsets and Cordless Handsets.

A handset typically has a keyboard and often a display. A handset can also be mobile or embedded device, where Skype is running inside the handset itself.

Examples of Handset audio UI devices are mobile phones and landline phones; few pictures below illustrate the group:

1.2.3 Speakerphone audio UI group

A Speakerphone audio UI product can be speakerphone, handset with speakerphone mode support or similar. Speakerphone audio UI product consists of two main components – microphone(s) and loudspeaker(s), usually integrated into the same device, but separate microphone and loudspeaker can also be viewed as a speakerphone. Often the device is placed on the table without physical contact with the user.

Copyright © 2009 Skype Inc. All Rights Reserved.

From audio quality perspective, quite crucial issue is a big enough distance between microphone and loudspeaker compared to distance to the user. This is due the need to achieve good acoustic echo cancellation from loudspeaker to microphone.

Unlike the headset and handset audio UI devices, the speakerphone audio UI device can be shared by several users, for example between users who sit around the table in a conference call. Conference calls are typically what speakerphones are used for. The speakerphone system may include several microphones or/and loudspeakers to enable picking up sound from all directions without attenuation and providing adequate sound volume to all conference call participants.

A speakerphone audio UI device is typically connected to the USB port or soundcard of a computer, but it can also be wireless. It can have keypad and display.

Speakerphone Skype certification specification is valid for Speakerphone audio UI products. Note that a handset or in principle even a headset can have a speakerphone audio UI functionality, and thus belong to Speakerphone audio UI group.

Examples of speakerphone audio UI devices are:

1.2.4 Other audio product group

This product is a part of audio signal chain in Skype environment, and it **does not provide acoustic user interface**, but still it can have a strong impact upon the audio quality for the end-to-end user experience. Typically it is an interface device that provides a conversion of audio from one format to another and thus does not improve the speech quality as such. These products can degrade the quality with additional delay, bandwidth limitation, noise, distortion, interference problems, etc.

The products belonging to this group are for example sound cards, Analog Terminal (Telephone) Adapters (ATA) and motherboards. As examples, here are an ATA device that turns common landline phone into a Skype internet phone and few soundcards:

1.2.5 Non-audio product group

Group contains products that actually do not directly influence audio, like cameras without microphone, displays, flash dongle... Such products can still have influence upon the audio quality, by increasing delay or creating drops or distortion of audio by overloading the computer or device in which the Skype application is running.

Below is an example of memory card that belongs to this group:

Copyright © 2009 Skype Inc. All Rights Reserved.

1.3 Audio requirements and priorities – overview

Audio requirements presented in this document aim at the products that provide a good sound quality, delight the user with great conversation experience and make communication easy.

At a high level the audio requirements and test cases in this document define the audio performance of a product. Some audio ergonomic requirements are set in other Skype Certification requirements.

The testing of audio quality is divided into objective and subjective testing. Objective testing measures quality by means of technical measurement tools, whereas subjective testing requires people to talk or/and listen and rate audio quality of the products. Audio performance requirements defined in this document are mainly verified using objective measures, but there are few cases where subjective measures are also involved.

1.3.1 Audio performance

The audio performance defines the audio quality of the product under test. In a high level the attributes that affect to the performance are intelligibility, naturalness and conversational effort. In a low level the performance consists of technical parameters such as frequency response, sensitivity, distortion, noise and acoustic echo.

Naturalness and also intelligibility are typically measured with listening quality metrics. Intelligibility can be difficult to measure, however a good assumption is that if user perceives the naturalness of conversation to be good then also the intelligibility must be good. Thus the listening quality metric that mainly concentrates to naturalness covers also enough of the intelligibility. The conversational quality metrics measure conversational effort.

1.3.2 Quality expectation of the audio UI groups

Audio quality expectations that the end user has for the product may vary depending on the price, advertisement promises and brand expectations, intended use of the product and experience of other similar solutions.

The audio requirements here are set based on the audio UI groups mainly, but in addition, there are a few technology dependent requirements. All requirements are the same for any product price category.

An example of technology dependency is cordless headsets technology limitation compared to plug-in headsets. Because of technology limitations the cordless headset like Bluetooth or DECT are often frequency band limited between 300 and 3.4 kHz (narrowband), like most landline and mobile phones are today. However Skype can provide wideband quality with frequencies between 50 and 7000 kHz. So Cordless headsets often can not benefit fully better audio quality, compared to the plug-in headsets, i.e. headsets with analog audio or USB connection, that do not have such limitation.

1.3.3 Use of the test case priorities

Each audio UI group has its own requirements and in addition there are General audio requirements valid for all groups in Chapter 2. The total number of test cases in for each solution varies between 10 and about 25. Each test case has several requirements and every requirement has a different priority.

The priorities are mapped to Must, Should, and Nice requirements.

They are marked as:

- Priority 1 = Must (at least 100% of Priority 1 requirements must PASS)
- Priority 2 = Should (at least 50% of Priority 2 requirements must PASS)
- Priority 3 = Nice to have (at least 10% of Priority 3 requirements must PASS)

Copyright © 2009 Skype Inc. All Rights Reserved.

2. General audio requirements valid for all groups

2.1 All groups: Audio performance requirements

Requirements below are valid for all groups: Headset, Handset, Speakerphone, Other and Non-audio products. Some of the requirements below are not applicable for Non-audio product. Audio test instructions in section 2.2 apply and should be followed in all requirements.

2.1.1 Priority: 1 Round trip delay of speech signals

Purpose: To ensure that both parties can hear each other without significant delay, the round trip acoustic end-to-end delay during Skype call must be as short as possible. When the delay is long the potential acoustic echo coming back to the talker is very disturbing. The interactivity of the interaction of call also suffers due to the long talk switching times between the call participants and there is a high risk of unintended doubletalk. The purpose of this test case is to ensure that the device under test does not increase the round trip delay in good network conditions over a specified limit.

Input: Play the measurement signal – first in sending and then in receiving direction. The delay is calculated using a cross correlation calculation. Short test signal is used for measuring delay at given moment. Long 60 second signal is used to determine the long term stability of the delay.

Round trip delay figure is calculated as Round trip delay = Sending direction delay + Receiving direction delay

Output: The average calculated round trip delay must be less than:

- 400ms – for devices connected to PC or MAC and using the software Skype client
- 400ms – for devices with embedded Skype client and using LAN cable
- 480ms – for wireless devices with embedded Skype client

Note: Please refer to 8.1.1 for description and specification of the measurement setup

2.1.2 Priority: 1 Total quality loss in sending direction

Purpose: To verify that users perceive natural and intelligible speech. The Perceptual Evaluation of Speech Quality tool (PESQ) [10] that complies with ITU-T P.862 standard is used for the analysis.

Input: Play back speech samples in sending direction (i.e. mic direction) and record the far end output.

Output: Use PESQ tool to analyze the speech quality in sending direction. Verify that the listening quality at the far end does not drop more than 1.0 MOS compared to a good quality reference device from the same product category measured in the same usage scenario.

If the device under test fails to meet the requirement the audio engineer will try to determine by listening to the recordings made during the above testing, if some of the following problems could be the cause for low MOS Listening Quality Objective (MOS-LQO) score:

- Speech quality is degraded by additional coding or format conversions

Copyright © 2009 Skype Inc. All Rights Reserved.

- Drops or distortions are present in speech signals
- Additional noises or sounds are present in speech signals
- Interference noises are present from electric power supply
- Interferences are present from devices with radio frequency transmission

Note: Skype wants to point out clearly that Skype acknowledges the fact that PESQ has not been designed and verified for acoustic interfaces therefore PESQ is not used as a measure of a quality of acoustic interface, but only to measure problems mentioned in the list up. Further Skype uses PESQ as a relative metric comparing the result of an acoustic interface device to a known reference device. In other words Skype is not using PESQ as an absolute metric in acoustic interface cases.

2.1.3 Priority: 1 Total quality loss in receiving direction

Purpose: To verify that users perceive natural and intelligible speech The Perceptual Evaluation of Speech Quality tool (PESQ) [10] that complies with ITU-T P.862 standard is used for the analysis.

Input: Play back speech samples in receiving direction (i.e. loudspeaker/earpiece direction) and record the near end output.

Output: Use PESQ tool to analyze the speech quality in receiving direction. Verify that the listening quality at the near end does not drop more than 1.0 MOS compared to a good quality reference device from the same product category measured in the same usage scenario.

If the device under test fails to meet the requirement the audio engineer will try to determine by listening to the recordings made during the above testing, if some of the following problems could be the cause for low MOS-LQO score:

- Speech quality is degraded by additional coding or format conversions
- Drops or distortions are present in speech signals
- Additional noises or sounds are present in speech signals
- Interference noises are present from electric power supply
- Interferences are present from devices with radio frequency transmission

Note: Skype wants to point out clearly that Skype acknowledges the fact that PESQ has not been designed and verified for acoustic interfaces therefore PESQ is not used as a measure of a quality of acoustic interface, but only to measure problems mentioned in the list up. Further Skype uses PESQ as a relative metric comparing the result of an acoustic interface device to a known reference device. In other words Skype is not using PESQ as an absolute metric in acoustic interface cases.

2.2 All groups: Additional requirements for PC or Mac accessories

2.2.1 Priority: 1 Analog gain adjustment latency

Purpose: To verify that the time to set- and get the microphone slider value does not exceed the requirement.

Input: Calculate the average time to set- and get the microphone slider value through Windows audio API.

Output: The average response time is < 50 ms

Copyright © 2009 Skype Inc. All Rights Reserved.

Note: Only applicable to devices using PC or Mac Skype Client.

2.2.2 Priority: 1 Device – Sampling frequency accuracy

Purpose: To ensure stable echo canceller performance the sampling frequencies of analog-to-digital and digital-to-analog converters must be accurate. This will allow using different audio interfaces for input and output during Skype call. For example: Using built-in speakers for Skype audio playback and USB microphone for Skype audio input.

Input: Measure the sampling frequencies at input and output when a sampling frequency of 48 kHz is selected. The sampling frequencies may be estimated by software using following calculation:

Fs(input) = number of samples recorded / measurement time
Fs(output) = number of samples played out / measurement time

The measurement time is >15 minutes and high precision timer is used. The number of samples being played out and recorded can be acquired through the audio API.

Output: Maximum deviation from the 48 kHz is 0.1%, i.e. 1000ppm for both play out and recording.

Note: Only applicable to devices using PC or Mac Skype Client.

2.3 General audio test instructions

Test environment is defined in Chapter 8.

There are good quality reference devices for each Audio UI groups separately. The reference device is chosen from the same Audio UI group from where the DUT is. Mean Opinion Scores and other audio performance measures from these devices are used as references for DUT.

2.3.1 Objective testing measurement setup

Audio testing tools and environment are listed in 8.1.1. Objective testing is performed with the automated audio testing system. Test practices and setups follow the principles given in ITU-T recommendations [4]. Actual test cases are specially built for the requirements defined in this document.

If Mean Opinion Score is mentioned in requirement, the result is judged by PESQ. Several test speech samples are recorded from sending and receiving directions. These recordings are divided to 10 sec length segments that are analyzed with objective speech quality tool. The speech material consists of variety of speakers and both male and female voices. The average score is used as the final MOS value.

In the test cases 2.1.2 – 2.1.3 MOS is first evaluated for a good quality reference device. Reference device belongs to the same audio UI group. Next the MOS is evaluated for DUT and the values are compared to each other. If the MOS value of DUT is lower than that of the reference device, then the audio engineer goes through the checklist and verifies which one of the conditions listed in the output of the test cases is not fulfilled causing the system to show lower MOS. This manual verification is performed both by listening to and analyzing the recordings.

If DUT has acoustic interface, the instructions from Sections 3.4, 4.3, and 5.3 will be followed for acoustic test setup.

The delay in the test case 2.1.1 is measured as follows:

- Skype call is created between two Skype clients.
- One Skype client runs on PC with Windows XP operating system (reference client).

Copyright © 2009 Skype Inc. All Rights Reserved.

- The other Skype client is run either on another PC or is embedded into device under test (referred to as device under test Skype client).
- A third computer with ACQUA audio measurement system, MFE front end and HATS connected to it is used that allows playback and recording simultaneously.
- A test signal is played at one end of a Skype-to-Skype call and recorded at the other end.
- The measurement signal is fed into the system either by electric connections or acoustically via the HATS mouth, depending on the test case.
- Delay measurements are performed in a local network with minimum number of clients on the same subnet.

Copyright © 2009 Skype Inc. All Rights Reserved.

3. Headset audio UI group

Audio test instructions in section 3.4 apply and should be followed in requirements of this Chapter.

3.1 Headset: Audio performance requirements

In all tests related to the requirements below the headset is positioned on HATS [2] as naturally as possible. HATS [2] is placed into the anechoic room.

3.1.1 Priority: 1 Microphone – Sensitivity at normal speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Play back a speech signal from the artificial mouth [2] at a normal speech level (check 3.4 Headset: Audio test instructions and Abbreviations). Microphone gain level is set by Skype client.

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -30 dBov RMS (-24 dBm0 RMS).

3.1.2 Priority: 2 Microphone – Sensitivity at lowered speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Play back a speech signal from the artificial mouth [2] at a lowered speech level (check 3.4 Headset: Audio test instructions and Abbreviations). Microphone gain level is set by Skype client.

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -30 dBov RMS (-24 dBm0 RMS).

3.1.3 Priority: 1 Microphone – Sensitivity at loud speech level

Purpose: To check that microphone circuit has enough dynamic headroom for occasions where loud speech level is used.

Input: Play back a speech signal from the artificial mouth [2] at a loud speech level (check 3.4 Headset: Audio test instructions and Abbreviations). Microphone gain level is set by Skype client.

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -30 dBov RMS (-24 dBm0 RMS). The signal must not overload the input causing clipping.

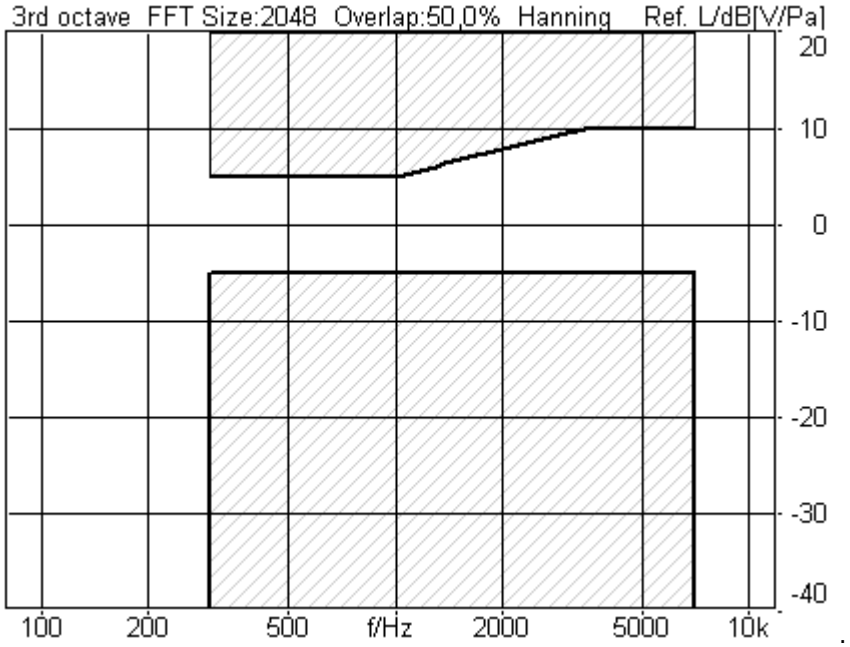
3.1.4 Priority: 1 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes minimum requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a limited wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.



Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
3400 Hz	-5,0 dB	10,0 dB
7000Hz	-5,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

Exception: In special cases an exception to this requirement can be given to products, where technology limits the bandwidth. Such cases can be DECT or Bluetooth products. The resulting frequency response in such cases must be at least 300 Hz – 3.4 kHz with a maximum ±5 dB ripple.

3.1.5 Priority: 2 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes super wideband requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.

Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
3400 Hz	-5,0 dB	10,0 dB
7000Hz	-5,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

3.1.6 Priority: 1 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 40 dB.

3.1.7 Priority: 2 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 45 dB.

Copyright © 2009 Skype Inc. All Rights Reserved.

3.1.8 Priority: 3 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 50 dB.

3.1.9 Priority: 2 Microphone – Speech to self noise ratio during speech activity

Purpose: To check that the self noise level of the microphone is sufficiently low during the active speech.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level. Immediately following play a special speech type of test signal to deactivate the possible microphone noise gating function. Record the test signal at the far end.

Output: The recorded microphone signal is processed to separate the speech part from the noise part. When the level of speech part is compared to the level of noise part, A-weighted RMS speech to noise ratio is at least 30 dB.

3.1.10 Priority: 2 Microphone – Speech to background noise ratio

Purpose: To verify that the microphone does not pick too much surrounding sounds and background noise compared to speech.

Input: Set up 3-dimensional sound playback environment into anechoic room. (Skype uses 18.1 channel 3D loudspeaker system using DIRAC processed samples). Remove HATS from the measurement area. Create different types of background noise environments to a measurement position, such as car, restaurant, street and office noises. Calibrate the A-weighted SPL level of noises to be 62 dB. Place HATS to the center of measurement area. Play back a measurement speech signal from the HATS artificial mouth [2] at a normal speech level and a background noise from the loudspeaker(s).

Output: The microphone signal is monitored at the far end output. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 10 dB.

3.1.11 Priority: 1 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 3.4 Headset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 40 dB.

3.1.12 Priority: 2 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 3.4 Headset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 45 dB.

Copyright © 2009 Skype Inc. All Rights Reserved.

3.1.13 Priority: 3 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 3.4 Headset: Audio test instructions and Abbreviations)

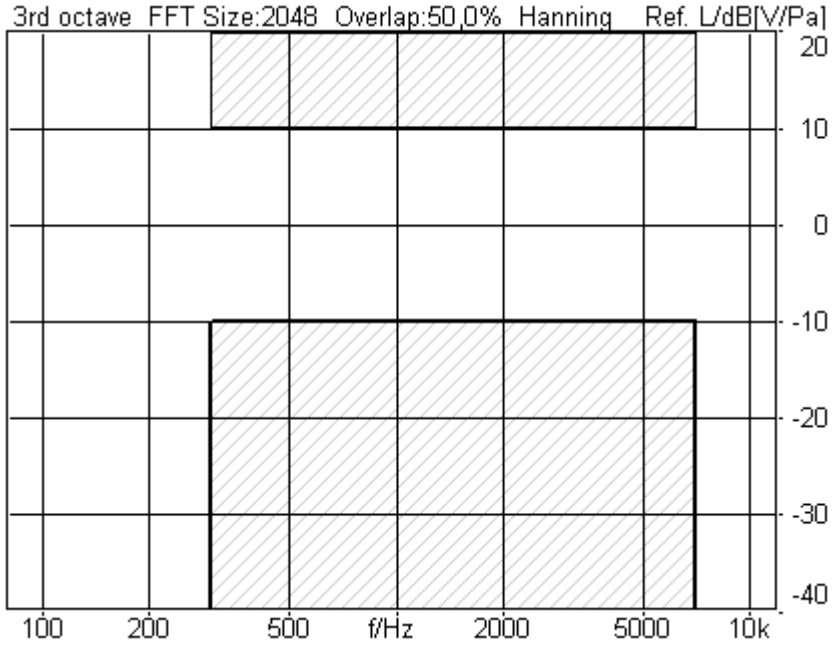
Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 50 dB.

3.1.14 Priority: 1 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes minimum requirement.

Input: Play a speech or a measurement signal through the earpiece.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a limited wideband tolerance window:



Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-10,0 dB	10,0 dB
7000Hz	-10,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

Exception: In special cases an exception to this requirement can be given to products, where technology limits the bandwidth. Such cases can be DECT or Bluetooth products. The resulting frequency response in such cases must be at least 300 Hz – 3.4 kHz with a maximum ±10 dB ripple.

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 3.4.1.

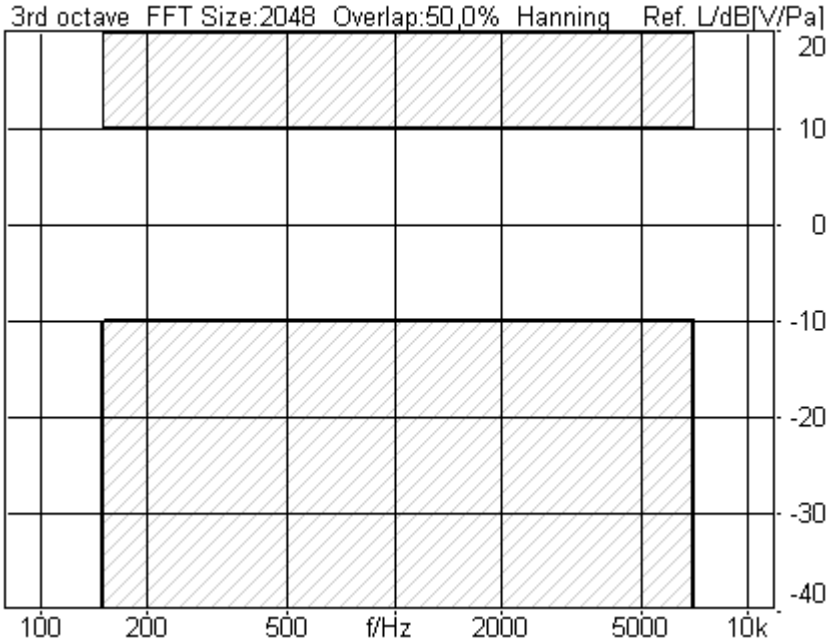
Copyright © 2009 Skype Inc. All Rights Reserved.

3.1.15 Priority: 2 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes super wideband requirement.

Input: Play a speech or a measurement signal through the earpiece.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a wideband tolerance window:



Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-10,0 dB	10,0 dB
7000Hz	-10,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 3.4.1.

3.1.16 Priority: 1 Earpiece – Stability of frequency response

Purpose: To check that frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece, Change the position of the headset on HATS and repeat the measurement several times.

Output: Compared to the normal position of the headset i.e. the frequency response got in the previous requirement, check if the maximum absolute change between 500 Hz and 1 kHz is less than 15 dB and between 1-3.4 kHz less than 10 dB.

Copyright © 2009 Skype Inc. All Rights Reserved.

3.1.17 Priority: 2 Earpiece – Stability of frequency response

Purpose: To check that the frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece, Change the position of the headset on HATS and repeat the measurement several times.

Output: Compared to the normal position of the headset i.e. the frequency response got in the previous requirement, check if the maximum absolute change between 300 and 1 kHz is less than 10 dB and between 1 kHz and 6 kHz less than 5 dB.

3.1.18 Priority: 3 Earpiece – Stability of frequency response

Purpose: To check that the frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece, Change the position of the headset on HATS and repeat the measurement several times.

Output: Compared to the normal position of the headset i.e. the frequency response got in the previous requirement, check if the maximum absolute change between 150 and 300 Hz is less than 10 dB and between 300 Hz and 7 kHz less than 5 dB.

3.1.19 Priority: 1 Minimum crosstalk from receiving to sending direction

Purpose: To check that crosstalk level between microphone and earpiece/loudspeaker meets the requirement. To ensure that conversation is pleasant and smooth, the echo must be minimized. Most of this echo is created between earpiece/ loudspeaker and microphone, but also electric connections and wires can leak i.e. to create crosstalk. This electric leaking is studied here.

Input: Cover microphone and/or earpiece/loudspeaker properly to minimize acoustic echo from earpiece/loudspeaker to microphone. Play back a test signal through device under test earpiece / loudspeaker. At the same time monitor and analyze the microphone signal level at the other Skype client output.

Output: Digital crosstalk level at the far end Skype client output is less than -51 dBov A-weighted RMS (-45 dBm0 A-weighted RMS).

3.2 Headset: requirements for Skype Super Wideband Certification (optional)

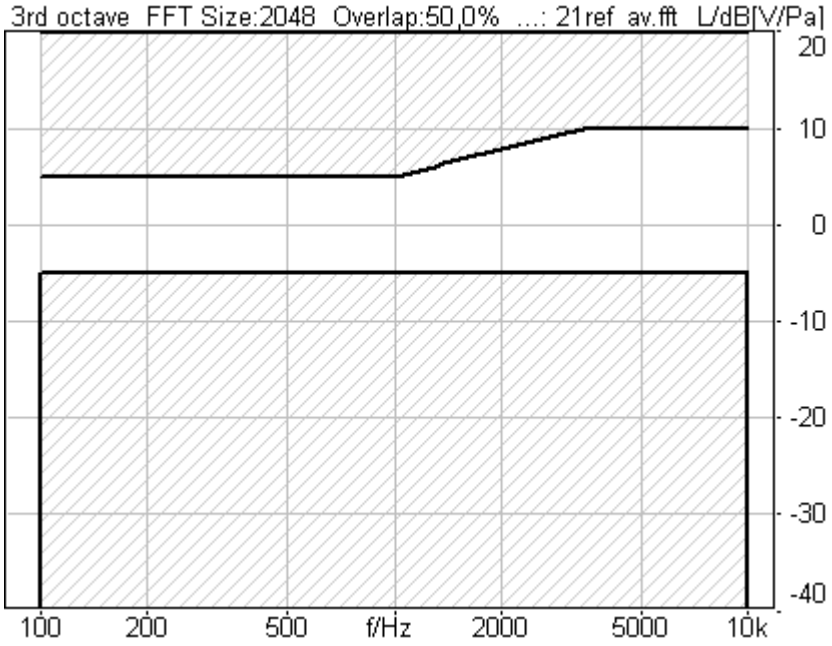
3.2.1 Priority: 1 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes super wideband requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a super wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.



Frequency	Lower limit	Upper limit
99Hz	-80,0 dB	20,0 dB
100Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
3400 Hz	-5,0 dB	10,0 dB
10000Hz	-5,0 dB	10,0 dB
10001Hz	-80,0 dB	20,0 dB

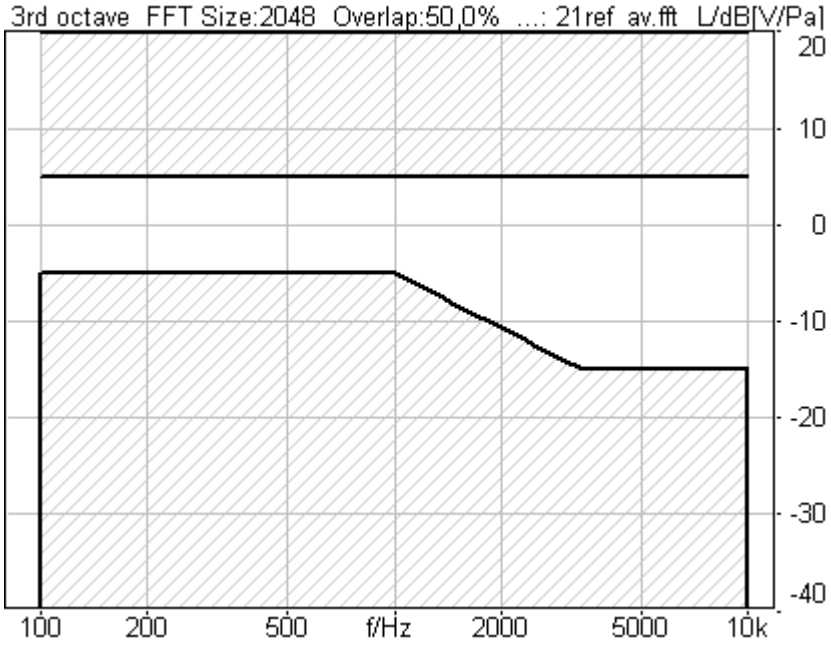
3.2.2 Priority: 1 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes super wideband requirement.

Input: Play a speech or a measurement signal through the earpiece.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a super wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.



Frequency	Lower limit	Upper limit
99Hz	-80,0 dB	20,0 dB
100Hz	-5,0 dB	5,0 dB
1000Hz	-5,0 dB	5,0 dB
3400Hz	-15,0 dB	5,0 dB
10000Hz	-15,0 dB	5,0 dB
10001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 3.4.1.

3.2.3 Priority: 1 Earpiece – Speech to noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 3.4 Headset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 55 dB.

3.3 Headset: Supporting audio documentation requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (for certification testing purposes). Such documentation contains engineering data and engineering test data for the product. Earpiece below means the acoustic output component for sound playback to the user's ear, for example the small loudspeaker.

Copyright © 2009 Skype Inc. All Rights Reserved.

3.3.1 Priority: 1 Verifying supporting documentation for Headset Audio UI group

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Output: DUT arrives with supporting audio documentation that contains information about:

- Active signal processing: yes/no, if yes then:
  - Acoustic echo cancellation in sending (i.e. mic) or/and in receiving (i.e. earpiece) directions: yes/no
  - Noise suppression in sending or/and in receiving directions: yes/no
  - Automated Gain Control in sending or/and in receiving directions: yes/no
- Microphone: Directionality/design principle
- Microphone: Frequency range (lowest and highest audible frequencies)
- Earpiece: Lowest and highest designed audible frequencies in intended use case

3.4 Headset: Audio test instructions

Test environment is defined in Chapter 8.

Headset under test is compared to a good quality reference headset. This reference headset is chosen from Skype Certified headsets. The sending (microphone) and receiving (earpiece/loudspeaker) parts might be chosen from two different headsets.

3.4.1 Objective testing measurement setup

Audio performance requirements are measured with objective measurement tools. The measurements will be performed with Head And Torso Simulator (HATS) [2] with type 3.3 anatomic ears [6] and with automated audio testing system in anechoic room. The audio testing tools and environment are listed in 8.1.1.Test practices and setups follow the principles given in ITU-T recommendations [4]. Actual test cases are specially built for the requirements defined in this document.

The measurements will be performed mainly with Skype to Skype call and DUT audio drivers. For passive headsets a reference soundcard is used as electric interface.

Frequency response results are averaged to 1/3 octave frequency resolution.

In microphone measurements the headset is attached to HATS as naturally as the user would do it in real life scenario.

In speech to background noise ratio measurement of microphone (3.1.10) various background noises are tested, like inside car, street and cafeteria noises. The background noises are real life 3D sound recordings that will be replayed with 3D loudspeaker setup of the test facility.

Earpiece/s in requirements 3.1.11 - 3.1.19 and 3.2.2 - 3.2.3 is/are measured with anatomic artificial ears, ITU-T type 3.3 that have measurement microphones at Drum Reference Point [6] and using DRP to diffuse field frequency response correction for measured earpiece responses. Note that Skype does not consider this to be the optimum response for a headset and does not assume flat diffuse field corrected response to be target in designing rather the diffuse field correction is chosen here for practical purposes. If manufacturer request or Certification tester considers appropriate some of the other standard corrections such as DRP to ERP (Ear Reference Point) or DRP to free field can be used or taken into account when interpreting the results.

Preferred listening level is defined to be 75 dB SPL A-weighted for a headset that reproduces speech only to one ear and 69 dB SPL A-weighted for a headset that plays speech to both ears.

Copyright © 2009 Skype Inc. All Rights Reserved.

The headset is adjusted manually with natural placement forces and effort to ear so that it gives good acoustic sealing for earpiece towards the ear. Measurement can be repeated for the same or for different pairs of headphones (if available during testing) if it seems to be appropriate for the headset. In stability measurements headset is taken off from the head and replaced again after each measurement. At least three different measurements are performed in this way.

The SPL level of normal speech in these tests is 62 dB SPL A-weighted at 1 m distance in front of mouth. The level is based on ITU-T recommendations of real and artificial mouth speaking levels. The lowered speech is about 10 dB quieter and loud speech is about 10 dB louder. Note that in real life the speaking levels vary more than 10 dB depending on speaker, distance between people having conversation and environment.

Copyright © 2009 Skype Inc. All Rights Reserved.

4. Handset audio UI group

Audio test instructions in section 4.3 apply and should be followed in all requirements.

4.1 Handset: Audio performance requirements

In all tests related to the requirements below the handset is positioned on HATS [2] as naturally as possible. HATS is placed into the anechoic room.

4.1.1 Priority: 1 Microphone – Sensitivity at normal speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Play back a speech signal from the artificial mouth [2] at a normal speech level (check 4.3 Handset: Audio test instructions and Abbreviations). Microphone gain level is controlled by the Skype audio engine.

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -30 dBov RMS (-24 dBm0 RMS).

4.1.2 Priority: 2 Microphone – Sensitivity at lowered speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Play back a speech signal from the artificial mouth [2] at a lowered level (check 4.3 Handset: Audio test instructions and Abbreviations). Microphone gain level is controlled by the Skype audio engine.

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -34 dBov RMS (-28 dBm0 RMS).

4.1.3 Priority: 1 Microphone – Sensitivity at loud speech level

Purpose: To check that microphone circuit has enough dynamic headroom for occasions where loud speech level is used.

Input: Play back a speech signal from the artificial mouth [2] at a loud speech level (4.3 Handset: Audio test instructions and Abbreviations). Microphone gain level is set by the Skype audio engine

Output: The microphone signal level is monitored at the far end and measured with ACQUA. The speech level is not less than -30 dBov RMS (-24 dBm0 RMS). The signal must not overload the input causing clipping.

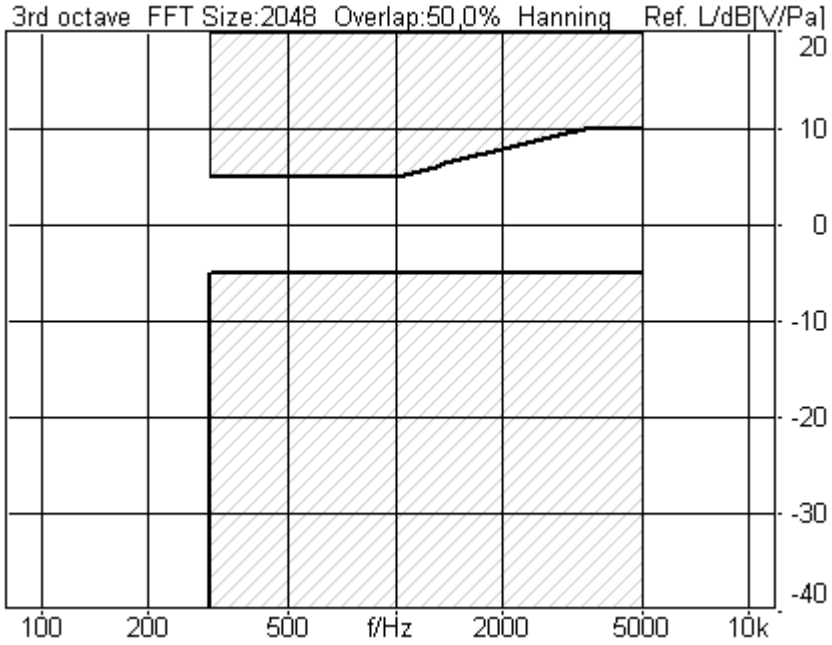
4.1.4 Priority: 1 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes minimum requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a limited wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.



Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
3400 Hz	-5,0 dB	10,0 dB
5000Hz	-5,0 dB	10,0 dB
5001Hz	-80,0 dB	20,0 dB

Note: Limited wideband is used because usually form factor of handsets does not allow positioning their microphones in front of the mouth and highest frequencies are attenuated due to the directionality of the mouth.

Exception: In special cases an exception to this requirement can be given to some cordless products. Such cases can be DECT or Bluetooth products in which the protocol limits the frequency bandwidth of speech. These are judged case by case.

4.1.5 Priority: 2 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes wideband requirement

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.

Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-5,0 dB	5,0 dB
1000 Hz	-5,0 dB	5,0 dB
3400 Hz	-5,0 dB	10,0 dB
7000Hz	-5,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

4.1.6 Priority: 1 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 40 dB.

4.1.7 Priority: 2 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 45 dB.

4.1.8 Priority: 3 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Copyright © 2009 Skype Inc. All Rights Reserved.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 50 dB.

4.1.9 Priority: 2 Microphone – Speech to self noise ratio during speech activity

Purpose: To check that the self noise level of the microphone is sufficiently low during the active speech.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level. Immediately following play a special speech type of test signal to deactivate the possible microphone noise gating function. Record the test signal at the far end.

Output: The recorded microphone signal is processed to separate the speech part from the noise part. When the level of speech part is compared to the level of noise part, A-weighted RMS speech to noise ratio is at least 30 dB.

4.1.10 Priority: 2 Microphone – Speech to background noise ratio

Purpose: To verify that the microphone does not pick too much surrounding sounds and background noise compared to speech.

Input: Set up 3-dimensional sound playback environment into anechoic room. (Skype uses 18.1 channel 3D loudspeaker system using DIRAC processed samples). Remove HATS from the measurement area. Create different types of background noise environments to a measurement position, such as car, restaurant, street and office noises. Calibrate the A-weighted SPL level of noises to be 62 dB. Place HATS to the center of measurement area. Play back a measurement speech signal from the HATS artificial mouth [2] at a normal speech level and a background noise from the loudspeaker(s).

Output: The microphone signal is monitored at the far end output. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 10 dB

4.1.11 Priority: 1 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 4.3 Handset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 40 dB.

4.1.12 Priority: 2 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 4.3 Handset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 45 dB.

4.1.13 Priority: 3 Earpiece – Speech to self noise ratio

Purpose: To check that the self noise level of the earpiece is sufficiently low.

Copyright © 2009 Skype Inc. All Rights Reserved.

Input: Play back a normal level speech signal at the far end input while on a Skype call. And adjust the listening level at near end output to the preferred listening level. (check 4.3 Handset: Audio test instructions and Abbreviations)

Output: The earpiece signal is monitored at the near end. When the speech signal level is compared to the noise level (noise is measured during pauses of the speech signal), A-weighted RMS speech to noise ratio is at least 50 dB.

4.1.14 Priority: 1 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes minimum requirement.

Input: Play a speech or a measurement signal through the earpiece of the handset.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a very limited wideband tolerance window

Frequency	Lower limit	Upper limit
499Hz	-80,0 dB	20,0 dB
500Hz	-10,0 dB	10,0 dB
5000 Hz	-10,0 dB	10,0 dB
5001Hz	-80,0 dB	20,0 dB

Exception: In special cases an exception to this requirement can be given to some cordless products. Such cases can be DECT or Bluetooth products in which the protocol limits the frequency bandwidth of speech. These are judged case by case.

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 4.3.1

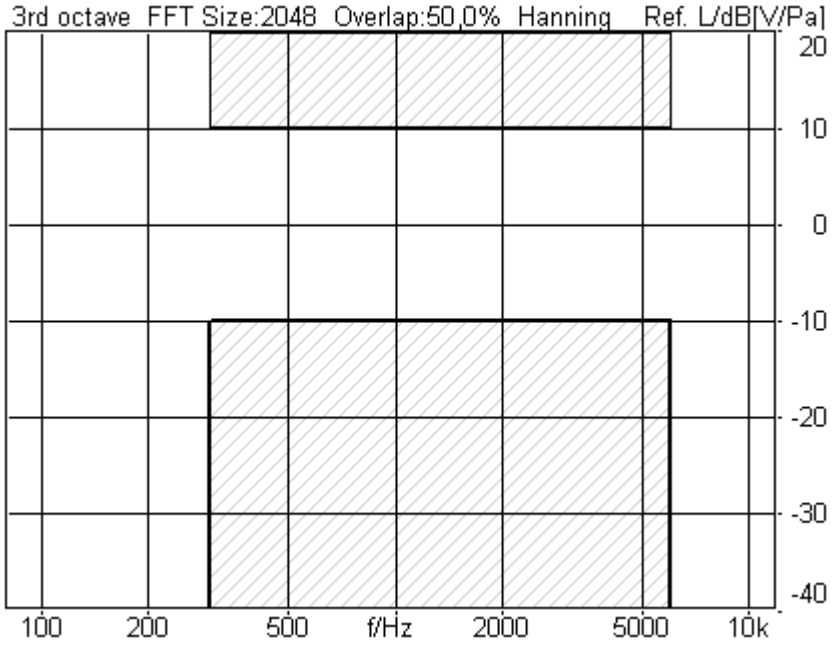
4.1.15 Priority: 2 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes limited wideband requirement.

Copyright © 2009 Skype Inc. All Rights Reserved.

Input: Play a speech or a measurement signal through the earpiece of the handset.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a limited wideband tolerance window:



Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-10,0 dB	10,0 dB
6000Hz	-10,0 dB	10,0 dB
6001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 4.3.1

4.1.16 Priority: 3 Earpiece – Frequency response

Purpose: To verify that the earpiece frequency response curve passes wideband requirement.

Input: Play a speech or a measurement signal through the earpiece of the handset.

Output: Measure frequency response of the earpiece by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.

Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-10,0 dB	10,0 dB
7000Hz	-10,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 4.3.1

4.1.17 Priority: 1 Minimum crosstalk from receiving to sending direction

Purpose: To check that crosstalk level between microphone and earpiece/loudspeaker meets the requirement. To ensure that conversation is pleasant and smooth, the echo must be minimized. Most of this echo is created between earpiece/ loudspeaker and microphone, but also electric connections and wires can leak i.e. to create crosstalk. This leakage is studied here.

Input: Cover microphone and/or earpiece / loudspeaker properly to minimize acoustic echo from earpiece/loudspeaker to microphone. Play back a test signal to device under test earpiece / loudspeaker. At the same time monitor and analyze the microphone signal level at the other Skype client output.

Output: Digital crosstalk level at other Skype client output is less than -51 dBov A-weighted RMS (-45 dBm0 A-weighted RMS).

4.1.18 Priority: 1 Earpiece – Stability of frequency response

Purpose: To check that the frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece. Change the position of handset on HATS and repeat the measurement several times.

Copyright © 2009 Skype Inc. All Rights Reserved.

Output: Compared to the normal position of the handset. Maximum absolute change between 500 Hz and 1 kHz is less than 15 dB and between 1-3.4 kHz less than 10 dB.

4.1.19 Priority: 2 Earpiece – Stability of frequency response

Purpose: To check that the frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece. Change the position of the handset on HATS and repeat the measurement several times.

Output: Compared to the normal position of the handset. Maximum absolute change between 300 and 1 kHz is less than 10 dB and between 1 kHz and 6 kHz less than 5 dB.

4.1.20 Priority: 3 Earpiece – Stability of frequency response

Purpose: To check that the frequency characteristic of the earpiece(s) does not change too much when its position on the ear changes, which can happen, when the user moves his head. Basically, this test case is to test leak tolerance of the earpiece.

Input: Play back a speech, music or measurement signal through the earpiece. Change the position of the handset on HATS and repeat the measurement several times.

Output: Compared to the normal position of the handset. Maximum absolute change between 150 and 300 Hz is less than 10 dB and between 300 Hz and 7 kHz less than 5 dB.

4.1.21 Priority: 1 Earpiece – Suitable volume level for office and home handset (Indoor)

Purpose: To verify that user can hear and understand speech while using the handset in normal every-day life.

Input: Play back speech through Skype and measure level with artificial ear (or listen the level subjectively).

Output: Earpiece volume level can be set both below and above 70 dB SPL A-weighted RMS (this is 5 dB below the preferred listening level).

4.1.22 Priority: 2 Earpiece – Suitable volume level for office and home handset (Indoor)

Purpose: To verify that user can hear and understand speech while using the handset in normal noisy office environment or home.

Input: Play back speech through Skype and measure level with artificial ear (or listen the level subjectively).

Output: Earpiece volume level can be set at least to 75 dB SPL A-weighted RMS (this is the preferred listening level).

4.1.23 Priority: 1 Earpiece – Suitable volume level for “anywhere” handset (Outdoor)

Purpose: To verify that user can hear and understand speech while using the handset in normal and noisy every-day-life environment.

Input: Play back speech through Skype and measure level with artificial ear (or listen the level subjectively).

Output: Earpiece volume level can be set at least to 75 dB SPL A-weighted RMS (this is the preferred listening level).

Note: “Anywhere” handset means a portable phone that can be used on street, public places, transportation, restaurants… where the environmental noise levels are high. Such a device can be mobile phones as an example.

Copyright © 2009 Skype Inc. All Rights Reserved.

4.1.24 Priority: 2 Earpiece – Suitable volume level for “anywhere” handset (Outdoor)

Purpose: To verify that user can hear and understand speech while using the handset in normal and noisy every-day-life environment.

Input: Play back speech through Skype and measure level with artificial ear (or listen the level subjectively).

Output: Earpiece volume level can be set at least to 80 dB SPL A-weighted RMS.

4.1.25 Priority: 1 Maximum ring tone loudness

Purpose: To verify that user can hear the ringing of incoming call in normal and noisy every-day-life environment.

Input: Set possible volume setting for ring tones of the handset at maximum. Set phone to the free field conditions (check Abbreviations for details). Choose calibrated measurement microphone and recording setup or take calibrated SPL measurement meter. Set the measurement microphone to 10 cm distance from the handset. The location can be chosen freely, but typically a place that gives highest SPL values is in the front of the outlets of the loudspeaker that plays the ring tones. Play ring tones of handset one by one.

Output: Record the ring tones with a calibrated microphone and check SPL levels offline or check SPL levels with SPL analyzer connected to the measurement microphone. For each tone measure the maximum hold value of SPL fast RMS (i.e. 125 ms exponential time weighting) with A-weighting (frequency correction) applied. The max-hold fast RMS level must be for half of the ring tones higher than 80 dB SPL.

Exception: Lower ring tone levels can be accepted if manufacturer asks for it with a detailed written explanation why and if Skype approves it. Such case can be for example that due to the design of the handset and safety regulations that are valid for the product, the highest output levels of the product needs to be limited to avoid too high sound pressure levels to user’s ears. One such regulation is given at European Standard EN 50332: “Sound system equipment: Headphone and earphones associated with portable audio equipment – Maximum sound pressure level measurement methodology and limit considerations”

4.1.26 Priority: 2 Maximum ring tone loudness

Purpose: To verify that user can hear the ringing of incoming call in normal and noisy every-day-life environment.

Input: Set possible volume setting for ring tones of the handset at maximum. Set phone to the free field conditions (check Abbreviations for details). Chose calibrated measurement microphone and recording setup or take calibrated SPL measurement meter. Set the measurement microphone to 10 cm distance from the handset. The location can be chosen freely, but typically a place that gives highest SPL values is in the front of the outlets of the loudspeaker that plays the ring tones. Play ring tones of handset one by one.

Output: Record the ring tones with a calibrated microphone and check SPL levels offline or check SPL levels with SPL analyzer connected to the measurement microphone. For each tone measure the maximum hold value of SPL fast RMS (i.e. 125 ms exponential time weighting) with A-weighting (frequency correction) applied. The max-hold fast RMS level must be for half of the ring tones higher than 90 dB SPL.

Exception: Lower ring tone levels can be accepted if manufacturer asks for it with a detailed written explanation why and if Skype approves it. Such case can be for example that due to the design of the handset and safety regulations that are valid for the product, the highest output levels of the product needs to be limited to avoid too high sound pressure levels to user’s ears. One such regulation is given at European Standard EN 50332: “Sound system equipment: Headphone and earphones

Copyright © 2009 Skype Inc. All Rights Reserved.

associated with portable audio equipment – Maximum sound pressure level measurement methodology and limit considerations"

4.1.27 Priority: 3 Maximum ring tone loudness

Purpose: To verify that user can hear the ringing of incoming call in normal and noisy every-day-life environment.

Input: Set possible volume setting for ring tones of the handset at maximum. Set phone to the free field conditions (check Abbreviations for details). Chose calibrated measurement microphone and recording setup or take calibrated SPL measurement meter. Set the measurement microphone to 10 cm distance from the handset. The location can be chosen freely, but typically a place that gives highest SPL values is in the front of the outlets of the loudspeaker that plays the ring tones. Play ring tones of handset one by one.

Output: Record the ring tones with a calibrated microphone and check SPL levels offline or check SPL levels with SPL analyzer connected to the measurement microphone. For each tone measure the maximum hold value of SPL fast RMS (i.e. 125 ms exponential time weighting) with A-weighting (frequency correction) applied. The max-hold fast RMS level must be for half of the ring tones higher than 100 dB SPL.

Exception: Lower ring tone levels can be accepted if manufacturer asks for it with a detailed written explanation why and if Skype approves it. Such case can be for example that due to the design of the handset and safety regulations that are valid for the product, the highest output levels of the product needs to be limited to avoid too high sound pressure levels to user's ears. One such regulation is given at European Standard EN 50332: "Sound system equipment: Headphone and earphones associated with portable audio equipment – Maximum sound pressure level measurement methodology and limit considerations"

4.2 Handset: Supporting audio documentation requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (only for certification testing purpose). Such documentation contains engineering data and engineering test data for the product.

Earpiece below means the acoustic output device, for example a small loudspeaker. Ring tone loudspeaker means the component that reproduces ring tones. In some devices it is the same component that reproduces speech, in others it is a separate element.

4.2.1 Priority: 1 Verifying supporting documentation for Handset audio

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Output: DUT arrives with supporting audio documentation that contains information about:

- Active signal processing: yes/no, if yes then:
  - Acoustic echo cancellation: yes/no, in sending (i.e. mic) or/and receiving (i.e. earpiece) directions.
  - Noise suppression: yes/no, in sending or/and receiving directions
  - Automated Gain Control: yes/no, in sending or/and receiving directions
  - Other: describe what, sending or/and receiving directions
- Microphone: Directionality/design principle
- Microphone: Frequency range (lowest and highest audible frequencies)
- Earpiece: Lowest and highest designed audible frequencies

Copyright © 2009 Skype Inc. All Rights Reserved.

- Earpiece: Designed acoustic SPL for receiving speech signal at the user's ear
- Ring tones: Number of tones
- Ring tones: Types of tones: MP3, Midi, wav/PCM etc…
- Ring tones from the loudspeaker: Maximum level of the loudest ring tone at 10 cm distance from the handset in the free field conditions (SPL, fast time weighting max hold)
- Ring tones from the loudspeaker: Level that half of the ring tones exceed when volume settings are on maximum. Measured at 10 cm distance from the handset in the free field conditions (SPL, fast time weighting max hold)

4.3 Handset: Audio test instructions

Test environment is defined in Chapter 8.

Handset under test is compared to a good quality reference handset. This reference handset is chosen from Certified Skype handsets. The sending (microphone) and receiving (earpiece/loudspeaker) parts might be chosen from two different handsets.

4.3.1 Objective testing measurement setup

Audio performance requirements are measured with objective measurement tools. The measurements will be performed with Head And Torso Simulator (HATS) [2] with type 3.3 anatomic ears [6] placing handset to the handset positioner and with an automated audio testing system in anechoic room. The audio testing tools and environment are listed in 8.1.1.Test practices and setups follow the principles given in ITU-T recommendations [4]. Actual test cases are specially built for the requirements defined in this document.

The measurements will be performed mainly in Skype call having all speech enhancement algorithms as they are by default in Skype and potential device audio drivers.

Frequency response results are averaged to 1/3 octave frequency resolution.

For majority of tests (expect the stability tests of earpiece) measurements the handset is placed on head as naturally as user would do it.

In speech to background noise ratio measurement of microphone (4.1.10) – various background noises can be tested, like inside car, street and cafeteria noises. The background noises are real life 3D sound recordings that will be replayed with 3D loudspeaker setup of the test facility.

Earpiece of handset in requirements 4.1.11 – 4.1.24 is measured with anatomic artificial ears, ITU-T type 3.3 and using a handset positioner available for ITU-T defined artificial heads [2] to place the handset to realistic and repeatable position. The handset will be placed to a position where phone is visually and acoustically sealed tightly to the ear but so that position is natural. In ITU-T recommendations so called ERP position (check [7], Annex E for 3.3 type ears) is typically proposed to be used. Unfortunately often this position does not give good and natural sealing for small modern handsets. Thus for many handset a better position compared to ERP position is to move the handset backwards (towards the back of head) 0.5-1 cm so that end of the handset seals to the pinna "hill" behind the ear canal entrance, press the handset closer to the ear by 0.5-1 cm and fold handset to natural position.

DRP to diffuse field frequency response correction is applied in earpiece measurements. This is contrary to ITU-T practice for narrowband phones to use DRP to ERP correction. Skype does not consider DRP to diffuse field correction to be the optimum response for a handset and does not assume flat diffuse field corrected response to be target in designing rather the diffuse field correction is chosen here for practical purposes. It can be noted that difference between DRP to diffuse field and DRP to ERP is relatively small compared to frequency limits given in the requirements. If manufacturer request or Certification tester considers appropriate some of the other standard corrections such as DRP to ERP (Ear Reference Point) or DRP to free field can be used or taken into account when interpreting the results.

Copyright © 2009 Skype Inc. All Rights Reserved.

Earpiece measurement can be repeated for the same or for different handsets (if available during testing) if it seems to be appropriate for the handset. In stability measurements handset position is adjusted with the handset positioner by 2-10 millimeters to potential directions where handset might be positioned or move during the real use. At least three different measurements are performed in this way.

Preferred listening level is defined to be 75 dB SPL A-weighted for a handset that reproduces speech only to one ear.

The SPL level of normal speech is defined here to be about 62 dB SPL A-weighted at 1 m distance in the front of the mouth. The level is based on ITU-T recommendations of real and artificial mouth speaking levels. The lowered speech is about 10 dB quieter and loud speech is about 10 dB louder. Note that in real life the speaking levels vary more than 10 dB depending on speaker, distance between people having conversation and environment.

Copyright © 2009 Skype Inc. All Rights Reserved.

5. Speakerphone audio UI group

In all tests related to the requirements below the solution with speakerphone audio functionality is positioned on the table in front of the HATS [2] in a natural way, as it would be done by the end user. Testing is done in the anechoic room. Audio test instructions in section 5.3 apply and should be followed in all requirements.

5.1 Speakerphone: Audio performance requirements

5.1.1 Priority: 1 Microphone – Sensitivity at normal speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Place the device under test to recommended usage position. Play back a speech signal from the artificial mouth [2] at a normal speech level (check 5.3 Speakerphone: Audio test instructions and Abbreviations). Microphone gain level is controlled by the Skype audio engine.

Output: The microphone signal level is monitored at the far end and measured with ACQUA [11]. The speech level is not less than -34 dBov RMS (-28 dBm0 RMS).

Note: Check the exact measurement setup and positions of DUT and HATS from 5.3.1.

5.1.2 Priority: 1 Microphone – Sensitivity at lowered speech level

Purpose: To check that the DUT microphone provides speech signal strong enough for the Skype audio engine.

Input: Place the device under test to recommended usage position. Play back a speech signal from the artificial mouth [2] at quiet speech level (check 5.3 Speakerphone: Audio test instructions and Abbreviations). Microphone gain level is controlled by the Skype audio engine.

Output: The microphone signal level is monitored at the far end and measured with ACQUA [11]. The speech level is not less than -34 dBov RMS (-28 dBm0 RMS).

5.1.3 Priority: 1 Microphone – Sensitivity at loud speech level

Purpose: To check that microphone circuit has enough dynamic headroom for occasions where loud speech level is used.

Input: Place the device under test to recommended usage position. Play back a speech signal from the artificial mouth [2] at loud speech level (check 5.3 Speakerphone: Audio test instructions and Abbreviations). Microphone gain level is controlled by the Skype audio engine.

Output: The microphone signal is monitored from another Skype client and measured with ACQUA [11]. The speech level is not less than -34 dBov RMS (-28 dBm0 RMS). The signal must not overload the input causing clipping.

5.1.4 Priority: 1 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes minimum requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a

Copyright © 2009 Skype Inc. All Rights Reserved.

limited wideband tolerance window:

Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-10,0 dB	10,0 dB
5000Hz	-10,0 dB	10,0 dB
5001Hz	-80,0 dB	20,0 dB

Exception: In special cases an exception to this requirement can be given to products, where for example echo cancellation technology limits the bandwidth. The resulting frequency response in such cases must be at least 300 Hz – 3.4 kHz with a maximum ±10 dB ripple.

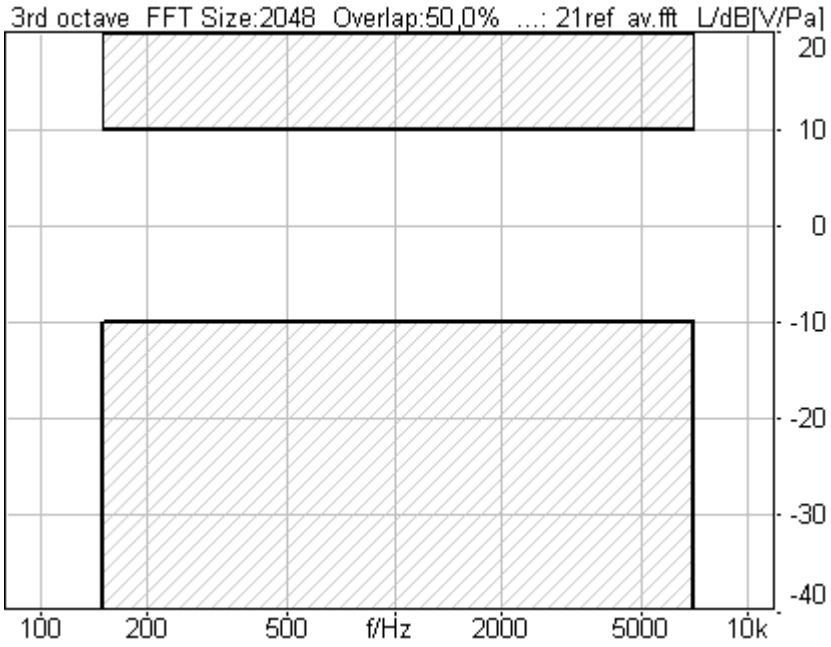
5.1.5 Priority: 2 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes wideband requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.



Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-10,0 dB	10,0 dB
7000Hz	-10,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

5.1.6 Priority: 3 Microphone – Frequency response

Purpose: To verify that the microphone frequency response curve passes super wideband requirement.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: Measure frequency response of the microphone by comparing the monitored speech signal to the original speech. The resulting frequency response fits into a super wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.

Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-10,0 dB	10,0 dB
10000Hz	-10,0 dB	10,0 dB
10001Hz	-80,0 dB	20,0 dB

5.1.7 Priority: 1 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Speakerphone is placed at the recommended usage distance from the HATS mouth. Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 35 dB.

5.1.8 Priority: 2 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Speakerphone is placed at the recommended usage distance from the HATS mouth. Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 40 dB.

Copyright © 2009 Skype Inc. All Rights Reserved.

5.1.9 Priority: 3 Microphone – Speech to self noise ratio

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Speakerphone is placed at the recommended usage distance from the HATS mouth. Play back a measurement signal from the artificial mouth [2] at a normal speech level to allow Skype to adjust the microphone gain setting to a suitable value. Then play the measurement signal again and record it at the far end.

Output: The recorded microphone signal is analyzed. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 45 dB.

5.1.10 Priority: 2 Microphone – Speech to self noise ratio during speech activity

Purpose: To check that the self noise level of the microphone is sufficiently low during the active speech.

Input: Play back a measurement signal from the artificial mouth [2] at a normal speech level. Immediately following play a special speech type of test signal to deactivate the possible microphone noise gating function. Record the test signal at the far end.

Output: The recorded microphone signal is processed to separate the speech part from the noise part. When the level of speech part is compared to the level of noise part, A-weighted RMS speech to noise ratio is at least 30 dB.

5.1.11 Priority: 1 Amount of acoustic echo

Purpose: To verify that the other party does not hear echo during the call. This is often the biggest problem for speakerphones.

Input: Ask another tester to use the product under test in a quiet meeting room (with the floor area of at least 10 m^2). Set the distance between him/her and the speakerphone to what the vendor specified as the recommended. If the recommended distance is not specified then use the maximum available distance. Pick yourself a good quality reference headset and set up a call. Ask the other party to set speakerphone volume level so that he/she could hear you with slight concentration effort, i.e. the audible speech level is about 5 dB below the preferred listening level (i.e. level here is about 55 dB SPL (A)). Try interrupting each other while talking.

Output: Talk and at the same time listen to the echo of your own voice which might come back to you from the other party. Echo may be audible, but if you don't speak at the same time the echo and other potential echo related artifacts are not annoying. Interrupting other party might be difficult as double talk transmission of system might be poor. Switching from talker to other is possible when there is a silence before talking turn changes.

Note: At Skype audio lab, the situation is simulated with an objective recording in call, where single talk from both parties, partial and heavy double talk are used. The recordings are analyzed by listening. The recordings and comments are included to the Audio report of the DUT.

Note 2: DUT is tested when Acoustic Echo Canceller of Skype is enabled.

5.1.12 Priority: 2 Amount of acoustic echo

Purpose: To verify that the other party does not hear echo during the call. This is often the biggest problem for speakerphones.

Input: Ask another tester to use the product under test in a quiet meeting room (with the floor area of at least 10 m^2). Set the distance between him/her and the speakerphone to what the vendor specified as the recommended. If the recommended distance is not specified then use the maximum available distance. Pick yourself a good quality reference headset and set up a call. Ask the other party

Copyright © 2009 Skype Inc. All Rights Reserved.

to set speakerphone volume level to the preferred listening level (60 dB SPL (A)). Try interrupting each other while talking.

Output: Talk and at the same time listen to the echo of your own voice which might come back to you from the other party. Echo may be audible when you talk at the same time, during this double talk the echo and other potential echo related artifacts are on maximum only slightly annoying. If you don't speak at the same time the echo is not audible and not annoying. Switching from one speaker to another with interruptions is easy.

Note: At Skype audio lab, the situation is simulated with an objective recording in call, where single talk from both parties, partial and heavy double talk are used. The recordings are analyzed by listening. The recordings and comments are included to the Audio report of the DUT.

Note 2: DUT is tested when Acoustic Echo Canceller of Skype is enabled.

Note 3: If there is Acoustic Echo Canceller in device, it will be tested also without Skype Acoustic Echo Canceller.

5.1.13 Priority: 3 Amount of acoustic echo

Purpose: To verify that the other party does not hear echo during the call. This is often the biggest problem for speakerphones.

Input: Ask another tester to use the product under test in a quiet meeting room (with the floor area of at least 10 m^2). Set the distance between him/her and the speakerphone to what the vendor specified as the recommended. If the recommended distance is not specified then use the maximum available distance. Pick yourself a good quality reference headset and set up a call. Ask the other party to set speakerphone volume level to the preferred listening level (60 dB SPL (A)). Try interrupting each other while talking.

Output: Talk and at the same time listen to the echo of your own voice which might come back to you from the other party. Echo is not audible and there are no other echo related artifacts. There are no challenges in switching from speaker to another with interruptions.

Note: At Skype audio lab, the situation is simulated with an objective recording in call, where single talk from both parties, partial and heavy double talk are used. The recordings are analyzed by listening. The recordings and comments are included to the Audio report of the DUT.

Note 2: DUT is tested when Acoustic Echo Canceller of Skype is enabled.

Note 3: If there is Acoustic Echo Canceller in device, it will be tested also without Skype Acoustic Echo Canceller.

5.1.14 Priority: 2 Echo loss in single talk during Skype call

Purpose: To verify that the other party does not hear acoustic echo while he or she is speaking.

Input: Place the speakerphone to the recommended usage position. A Skype call is set up and a test signal is sent to the loudspeaker. The output signal at the remote end is measured and analyzed for echo presence. Levels and details are defined in ITU-T Recommendations G.122 and G.131 [8], [9]

Output: Echo loss must be at least 40 dB.

Note: This requirement is based on ITU-T Recommendation G.131 [9] information of tolerance of talker echo (figure 1 in G.131) targeting to 50 dB Talker Echo Loudness Rating that corresponds to about 40 dB Echo Loss taking into account of potential long delay of IP network. The measurement method is defined ITU-T

Copyright © 2009 Skype Inc. All Rights Reserved.

Recommendation G.122 [8]. In this requirement the Skype speech preprocessing, including echo canceller is enabled.

5.1.15 Priority: 3 Echo loss in single talk without Skype speech improvements

Purpose: To verify that the other party does not hear acoustic echo while he or she is speaking.

Input: Place the speakerphone to the recommended usage position. A test signal is sent to the loudspeaker. The microphone output is measured and analyzed for echo presence. Levels and details are defined in ITU-T Recommendations G.122 and G.131 [8], [9]. In practice the loudspeaker and microphone levels are set to the same as they were in requirement 5.1.14.

Output: Echo loss must be at least 40 dB.

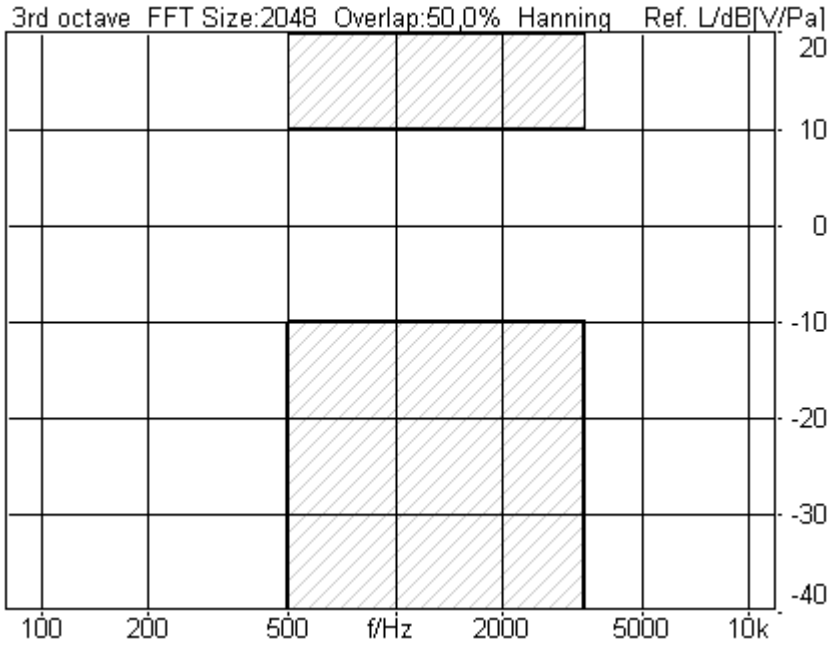
Note: This requirement is based on ITU-T Recommendation G.131 [9] information of tolerance of talker echo (figure 1 in G.131) targeting to 50 dB Talker Echo Loudness Rating that corresponds to about 40 dB Echo Loss taking into account of potential long delay of IP network. The measurement method is defined ITU-T Recommendation G.122 [8]. The measurement can be made during a Skype call, but having Skype echo canceller disabled.

5.1.16 Priority: 1 Loudspeaker – Frequency response

Purpose: To verify that the loudspeaker frequency response curve passes minimum requirement.

Input: Play back a measurement signal through the loudspeaker. Measure the loudspeaker frequency response at the recommended usage position.

Output: The resulting frequency response fits into a limited narrowband tolerance window:



Frequency	Lower limit	Upper limit
499Hz	-80,0 dB	20,0 dB
500Hz	-10,0 dB	10,0 dB
3400Hz	-10,0 dB	10,0 dB

Copyright © 2009 Skype Inc. All Rights Reserved.

3401Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 5.3.1

5.1.17 Priority: 2 Loudspeaker – Frequency response

Purpose: To verify that the loudspeaker frequency response curve passes wideband requirement.

Input: Play back a measurement signal through the loudspeaker. Measure the loudspeaker frequency response at the recommended usage position.

Output: The resulting frequency response fits into a limited wideband tolerance window:

Frequency	Lower limit	Upper limit
299Hz	-80,0 dB	20,0 dB
300Hz	-10,0 dB	10,0 dB
7000Hz	-10,0 dB	10,0 dB
7001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 5.3.1.

5.1.18 Priority: 3 Loudspeaker – Frequency response

Purpose: To verify that the loudspeaker frequency response curve passes super wideband requirement.

Input: Play a speech or a measurement signal through the loudspeaker. Measure the loudspeaker frequency response at the recommended usage position.

Output: The resulting frequency response fits into a limited super wideband tolerance window:

Copyright © 2009 Skype Inc. All Rights Reserved.

Frequency	Lower limit	Upper limit
149Hz	-80,0 dB	20,0 dB
150Hz	-10,0 dB	10,0 dB
10000Hz	-10,0 dB	10,0 dB
10001Hz	-80,0 dB	20,0 dB

Note: Skype uses ITU-T type 3.3 ear and DRP to diffuse field correction, check test instructions 5.3.1.

5.1.19 Priority: 1 Loudspeaker – Suitable volume level for quiet office use

Purpose: To verify that the speakerphone output level fulfills requirement for recommended operating distance in quiet office environment.

Input: Place the speakerphone at the recommended operating distance from the measurement microphone. Play back a speech signal through the loudspeaker. Set the speakerphone volume to loud and measure the loudspeaker output.

Output: Measured output is at least 55 dB SPL A-weighted (this is 5 dB below the preferred listening level).

5.1.20 Priority: 1 Loudspeaker – Distortion at quiet office use

Purpose: To verify that device does not create too much distortion to degrade speech quality and produce audible echo at the far end.

Input: Place the speakerphone to the recommended usage position. Make a Skype call and play a speech signal from other party side. Set the speakerphone volume to 55 dB SPL A-weighted. THD is then measured with stepped sine signals set to -9 dBov RMS (-15 dBm0 RMS) at other Skype client (equals to about -6 dBov peak level). Distortion is measured starting from the lowest corner frequency of cases 5.1.16-5.1.18 that DUT passed and up to 3.4 kHz. Speech level and distortion are measured with measurement microphone which is placed at the intended position of listener's head.

Copyright © 2009 Skype Inc. All Rights Reserved.

Output: Measured Total Harmonic Distortion (THD) of all measurement points is below 3% (equals to -30 dB).

Example: How to calculate the measurement bandwidth? For example if DUT passed cases 5.1.16 and 5.1.17, which have low corner frequencies of 500 and 300 Hz respectively, then the lower one of these is chosen, meaning that distortion will be measured from 300 to 3400 Hz.

5.1.21 Priority: 2 Loudspeaker – Suitable volume level for normal office use

Purpose: To verify that the speakerphone output level fulfills requirement for recommended operating distance in normal office environment.

Input: Place the speakerphone at the recommended operating distance from the measurement microphone. Play back a speech signal through the loudspeaker. Set the speakerphone volume to loud and measure its loudspeaker output.

Output: Measured output is at least 60 dB SPL A-weighted.

5.1.22 Priority: 2 Loudspeaker – Distortion at normal office use

Purpose: To verify that device does not create too much distortion to degrade speech quality and produce audible echo at the far end.

Input: Place the speakerphone to the recommended usage position. Make a Skype call and play a speech signal from other party side. Set the speakerphone volume to 60 dB SPL A-weighted. THD is then measured with stepped sine signals set to -9 dBov RMS (-15 dBm0 RMS) at other Skype client (equals to about -6 dBov peak level). Distortion is measured starting from the lowest corner frequency of cases 5.1.16-5.1.18 that DUT passed and up to 3.4 kHz. Speech level and distortion are measured with measurement microphone which is placed at the intended position of listener's head.

Output: Measured Total Harmonic Distortion (THD) of all measurement points is below 3% (equals to -30 dB).

5.1.23 Priority: 3 Loudspeaker – Suitable volume level for noisy office use

Purpose: To verify that the speakerphone output level fulfills requirement for recommended operating distance in noisy office environment.

Input: Place the speakerphone at the recommended operating distance from the measurement microphone. Play back a speech signal through the loudspeaker. Set the speakerphone volume to loud and measure the loudspeaker output.

Output: Measured output is at least 65 dB SPL A-weighted.

5.1.24 Priority: 3 Loudspeaker – Distortion at noisy office use

Purpose: To verify that device does not create too much distortion to degrade speech quality and produce audible echo at the far end.

Input: Place the speakerphone to the recommended usage position. Make a Skype call and play a speech signal from other party side. Set the speakerphone volume to 65 dB SPL A-weighted. THD is then measured with stepped sine signals set to -9 dBov RMS (-15 dBm0 RMS) at other Skype client (equals to about -6 dBov peak level). Distortion is measured starting from the lowest corner frequency of cases 5.1.16-5.1.18 that DUT passed and up to 3.4 kHz. Speech level and distortion are measured with measurement microphone which is placed at the intended position of listener's head.

Output: Measured Total Harmonic Distortion (THD) of all measurement points is below 3% (equals to -30 dB).

Copyright © 2009 Skype Inc. All Rights Reserved.

5.1.25 Priority: 2 Loudspeaker – Volume level at maximum operating distance

Purpose: To verify that the speakerphone output level fulfills requirement for recommended maximum operating distance in the office environment.

Input: Place the speakerphone at the recommended maximum operating distance from the measurement microphone. Play back a speech signal through the loudspeaker. Set the speakerphone volume to loud and measure the loudspeaker output.

Output: Measured output is at least 55 dB SPL A-weighted (this is 5 dB below the preferred listening level).

5.1.26 Priority: 2 Microphone – Sensitivity at maximum operating distance

Purpose: To verify that the DUT microphone provides strong speech signal for Skype application, when speakerphone is tested at the maximum operating distance that has been specified by he manufacturer. The distance is measured between the microphone and the mouth.

Input: Place the speakerphone to the maximum operating distance. Play back a speech signal from an artificial mouth [2] at a normal speech level from intended usage distance of the speakerphone.

Output: The microphone signal is monitored from another Skype client and measured. The speech level is not less than -34 dBov RMS (-28 dBm0 RMS).

5.1.27 Priority: 3 Microphone – Speech to self noise ratio at maximum operating distance

Purpose: To check that the self noise level of the microphone is sufficiently low.

Input: Place the speakerphone to the recommended usage position. Play back a measurement signal from the artificial mouth [2] at a normal speech level.

Output: The microphone signal is monitored at the far end and measured with ACQUA [11]. When the speech signal level is compared to the noise level (noise is measured during pauses of speech), A-weighted RMS speech to noise ratio is at least 35 dB.

5.2 Speakerphone: Supporting audio documentation requirements

In addition to the user manual (the one that comes with the product) in Certification testing we also ask for supporting audio documentation. Such documentation contains engineering data and engineering test data of the product.

5.2.1 Priority: 1 Verifying supporting documentation for Speakerphone audio

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Output: DUT arrives with supporting audio documentation that contains the following information:

- Usage related info:
  - Recommended operating distance
  - Maximum operating distance
- Active signal processing: yes/no, if yes then:
  - Active beam forming microphone and/or loudspeaker: yes/no
  - In built acoustic echo cancellation: yes/no
  - Echo cancellation operating bandwidth (narrowband, wideband, super wideband)
  - Noise suppression: yes/no, in sending or/and receiving directions

Copyright © 2009 Skype Inc. All Rights Reserved.

- Automated Gain Control: yes/no, in sending or/and receiving directions
- Other: describe what, sending or/and receiving directions

- Microphone/s:
  - Frequency range (lowest and highest audible frequencies)
  - Directionality/design principle of a microphone
  - Number of microphones / microphone inputs
  - Microphone phantom power yes/no, supply voltage (if applicable)
  - Microphone input connector type (balanced, unbalanced) (if applicable)
- Loudspeaker/s:
  - Frequency range (lowest and highest audible frequencies)
  - Number of loudspeaker / line outputs (if applicable)
  - Loudspeaker design principle (one/multiway, open/closed box/bass reflex)

5.3 Speakerphone: Audio test instructions

Test environment is defined in Chapter 8.

Device under test that provides Speakerphone acoustic UI is compared to a good quality reference speakerphone. This reference speakerphone is chosen from Skype Certified speakerphones. The sending (microphone) and receiving (loudspeaker) parts might be chosen from two different products.

5.3.1 Objective testing measurement setup

Audio performance requirements are measured with objective measurement tools. The measurements will be performed with Head And Torso Simulator (HATS) [2] or/and measurement microphone and with automated audio testing system. The measurements are performed in anechoic and/or in quiet office room. The audio testing tools and environment are listed in 8.1.1.Test practices and setups follow the principles given in ITU-T recommendations [4]. Actual test cases are specially built for the requirements defined in this document.

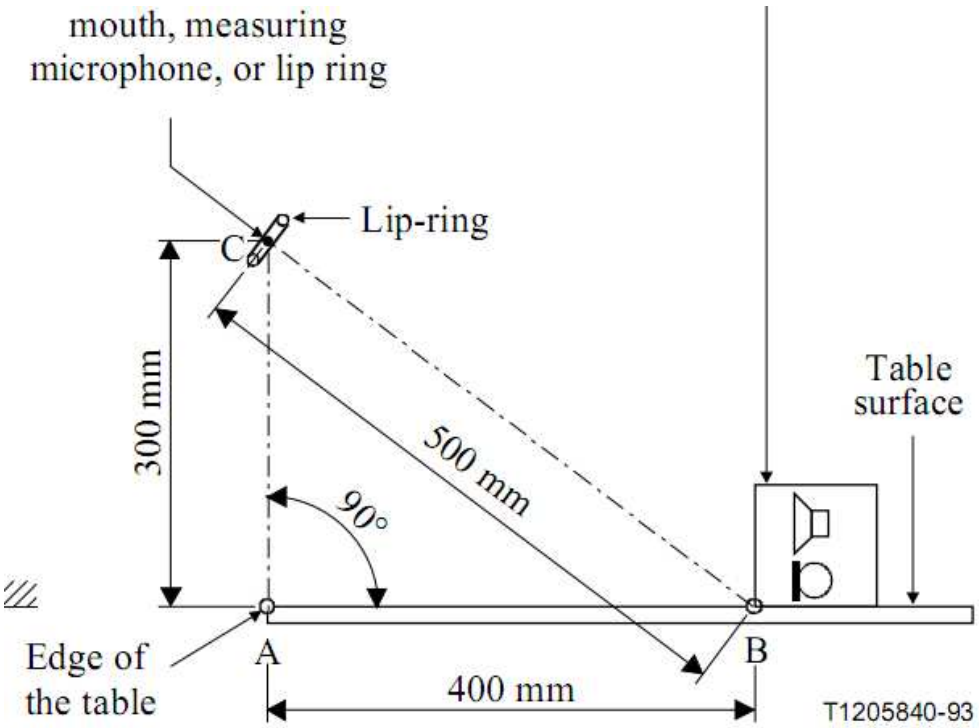
The measurements will be performed mainly during a Skype to Skype call. If device is connected to PC the default audio drivers for DUT are used.

Frequency response results are averaged to 1/3 octave frequency resolution.

For hand-held speakerphones (such as small phones) the recommended usage position is in front of the user -30˚ below the mouth at a recommended usage distance, specified by the manufacturer.

For non-handheld i.e. desktop devices, recommended usage position specified by the manufacturer shall be used. If there is no manufacturer's recommendation provided, the test arrangement as per ITU-T recommendation P.340 [5] shall be used (see figure below).

Copyright © 2009 Skype Inc. All Rights Reserved.



Preferred listening level is defined to be 60 dB SPL A-weighted RMS for a speakerphone.

The SPL level of normal speech in these tests is 62 dB SPL A-weighted at 1 m distance in the front of mouth. The level is based on ITU-T recommendations of real and artificial mouth speaking levels. The lowered speech is about 10 dB quieter and loud speech is about 10 dB louder. Note that in real life the speaking levels vary more than 10 dB depending on speaker, distance between people having conversation and environment.

5.3.2 Subjective testing measurement setup

A speakerphone is tested by a tester under normal conditions as defined in the requirements. The speakerphone is placed on the table next to the user in the office space or/and in a meeting room (with the floor area of at least 10 m2).

If the speakerphone is not a standalone product, then it is tested in a typical environment, for example on a PC with a normal sound card and audio connectors. The speakerphone will be tested on several PCs or/and operation systems if necessary.

The testing positions are set such as defined in Objective testing measurement setup in the previous section. The tester on the other side should use a good quality headset.

Subjective testing is applied primarily for Echo requirements 5.1.11 – 5.1.13. First create a Skype call and play/speak test signal from the other party side. The both sides of call should talk at normal levels. The volume level of speakerphone loudspeaker is set as is defined in the requirements - either to a lowered speech or preferred speech level. It is recommended that call is recorded and listened to detect potential echo.

During the test call testers are recommended to talk both single and double talk. At speakerphone side a tester can talk partly at the same time as the other party is speaking.

Very important issues with a speakerphone are the distance from the user and hard surfaces proximity, such as wall or computers. The hard surfaces create strong reflections and acoustic echo back to the microphone. This can be tested by using the device at the maximum distance and placing it closer to the wall(s) or computer(s).

Test case judgments are based on comparison of tester's perception with the requirements. The recorded samples can also be analyzed with a normal sound editor program available for

Copyright © 2009 Skype Inc. All Rights Reserved.

computers, check Section 8.1.1. Testing is by nature informal meaning that it does not have blind testing of multiple people and related statistical analysis of judgments. If judgment of some requirement is difficult, then two additional testers will perform the test case, and if two or more testers judge the case to be failed, then requirement is failed.

Copyright © 2009 Skype Inc. All Rights Reserved.

6. Other audio product group

Audio test instructions in section 6.3 apply and should be followed in all requirements.

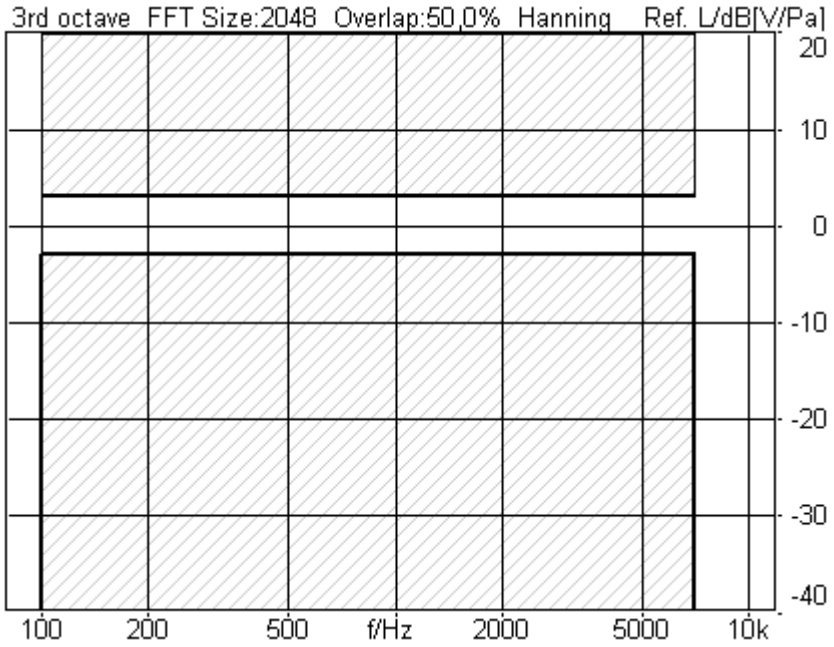
6.1 Other audio product: Audio performance requirements

6.1.1 Priority: 1 Frequency responses – sending and receiving directions

Purpose: To verify that the sending and receiving direction frequency response curves pass the minimum requirements.

Input: Play back a measurement signal in sending and receiving directions.

Output: Measure frequency responses of sending and receiving directions by comparing the monitored speech signals to the original speech signals. The resulting frequency responses fit into a wideband tolerance window:



Frequency	Lower limit	Upper limit
99Hz	-80,0 dB	20,0 dB
100Hz	-3,0 dB	3,0 dB
7000Hz	-3,0 dB	3,0 dB
7001Hz	-80,0 dB	20,0 dB

Exception: In special cases an exception to this requirement can be given for some cordless and Analog Telephony Adapter (ATA) products, like DECT or Bluetooth products in which the protocol limits the frequency bandwidth. These are judged case by case.

Copyright © 2009 Skype Inc. All Rights Reserved.

6.1.2 Priority: 1 Product provides suitable levels for audio signal output

Purpose: To verify that the output level that the product provides is suitable for other devices in the signal chain. For example a sound card needs to provide suitable (high enough) signal levels for reference headphones and ATA must provide suitable output levels for a reference handset.

Input: Using the product set up a Skype call. Depending on the product interface, connect corresponding reference device:

• Sennheiser HD650 Headphones
• Siemens Euroset 802 Deskset

Perform the level measurement test.

Output: Output volume level is at least 70 dB SPL A-weighted RMS (this is 5 dB below the preferred listening level on one ear listening case)..

6.1.3 Priority: 1 Product provides suitable levels for audio signal input

Purpose: To verify that the input level that the product provides is suitable for other devices in the signal chain. For example a sound card needs to provide suitable (high enough) signal levels for reference microphone and ATA must provide suitable input levels for a reference handset.

Input: Using the product set up a Skype call. Depending on the product interface, connect corresponding reference device:

• Microphone EMM-8
• Siemens Euroset 802 Deskset

Perform the level measurement test.

Output: Input volume level is not less than -30 dBov RMS (-24 dBm0 RMS).

6.1.4 Priority: 1 Minimum crosstalk from receiving to sending direction

Purpose: To verify if crosstalk level passes the minimum requirement.

Input: Disconnect the acoustic interface from the device. Play back a test signal to device under test output i.e. receiving direction. At the same time monitor and analyze the input i.e. sending direction signal level at the other Skype client output.

Output: Digital crosstalk level at other Skype client output is less than -51 dBov A-weighted RMS (-45 dBm0 A-weighted RMS).

6.2 Other audio product: Supporting audio documentation requirements

In addition to the user manual (the one that comes with the product) we also ask for supporting audio documentation (for certification testing purposes). Such documentation contains engineering data and engineering test data for the product.

6.2.1 Priority: 1 Verifying supporting documentation for Other audio product

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Output: DUT arrives with supporting audio documentation that contains the following information:

- Sending:
    - Speech signal delay from input to output (if above 5 ms)

Copyright © 2009 Skype Inc. All Rights Reserved.

    - Usable frequency bandwidth
- Receiving:
    - Speech signal delay from input to output (if above 5 ms)
    - Usable frequency bandwidth
- Connectors (if applicable):
    - Type/s & electric connections (ground, signals, bias voltages…)
    - Target input and output levels/voltages
    - Maximum input and output levels/voltages
    - Maximum and minimum impedances for external connection
- Volume control (if applicable):
    - Range in dB
    - Minimum volume (dBV, dBSPL or similar RMS)
    - Maximum volume (dBV, dBSPL or similar RMS)
- Active signal processing: yes/no
    - if yes then what?

6.3 Other audio product: Audio test instructions

6.3.1 Objective testing measurement setup

Objective testing arrangement depends on if the testing is performed with or without acoustic interface. Example of the earlier is a sound card that can be tested together with headset. The example of latter is an audio processing algorithm that does not give direct signal to acoustic interface device. In a case the device is tested together with acoustic interface, the testing setup can be picked from headset, handset or speakerphone test instructions in the previous chapters. In another case when acoustic interface is not used, the electric to electric tests between two Skype clients and the DUT can be performed.

The measurements will be performed mainly in Skype call having all speech enhancement algorithms as they are by default in Skype and potential device audio drivers.

Frequency response results are averaged to 1/3 octave frequency resolution.

Copyright © 2009 Skype Inc. All Rights Reserved.

7. Non-audio product group

Audio test instructions in section 7.3 apply and should be followed in all requirements.

7.1 Non-audio product: Audio performance requirements

7.1.1 Priority: 1 Continuous transmission of speech

Purpose: To verify that users hear continuous audio transmission without short or long temporal drops or distortions while product or solution is under normal and heavy load.

Input: Connect the product to the Skype or in between Skype signal path. Set up a call and play back speech samples in sending and receiving directions. Record both near and far end Skype outputs. During the recording create load to the product, solution or/and PC or phone to where Skype application is installed. Use for example other available Skype features, such as file sharing and video, open browser and open a video playback etc.

Output: Use PESQ tool to analyze the speech quality in both sending and receiving directions. The biggest MOS-LQO drop must be smaller than 1.0 compared to average MOS score of Skype call without the product in use over periods of 10 secs.

7.1.2 Priority: 2 Continuous transmission of speech

Purpose: To verify that users hear continuous audio transmission without short or long temporal drops or distortions while product or solution is under normal and heavy load.

Input: Connect the product to the Skype or in between Skype signal path. Set up a call and play back speech samples in sending and receiving directions. Record both near and far end Skype outputs. During the recording create load to the product, solution or/and PC or phone to where Skype application is installed. Use for example other available Skype features, such as file sharing and video, open browser and open a video playback etc.

Output: Use PESQ tool to analyze the speech quality in both sending and receiving directions. The biggest MOS-LQO drop must be smaller than 0.5 compared to average MOS score of Skype call without the product in use over periods of 10 secs.

7.2 Non-audio product: Supporting audio documentation

In addition to the user manual (the one that comes with the product) in Certification testing we also ask for supporting audio documentation. Such documentation contains engineering data and engineering test data of the product.

7.2.1 Priority: 1 Verifying supporting documentation for Non-audio product

Purpose: Solution must come with a supporting audio documentation (only for certification testing purposes).

Output: DUT arrives with supporting audio documentation that contains the following information:

- Solution speech signal delay (if above 5 ms)

Copyright © 2009 Skype Inc. All Rights Reserved.

- Influence on the audio quality of Skype call under normal and heavy loading of product/solution

7.3 Non audio product: Audio test instructions

7.3.1 Objective testing measurement setup

Audio performance requirements can be measured with objective measurement tools. Measurement tool and Skype clients will be connected electrically as acoustic interface is not needed.(electric to electric measurement).

Mean Opinion Score results are judged by PESQ or similar advanced objective speech quality metric. Several test speech samples are recorded from sending and receiving directions. These recordings are divided to about 10 sec length segments that are analyzed with objective speech quality tool. The speech material consists of variety of speakers and both male and female voices.

MOS values are calculated without and with the DUT being connected. The drop is calculated as a difference between individual MOS values.

Copyright © 2009 Skype Inc. All Rights Reserved.

8. List of environments

8.1 List of Test Platforms

Solutions are tested in the following environments. This list will be extended in future:

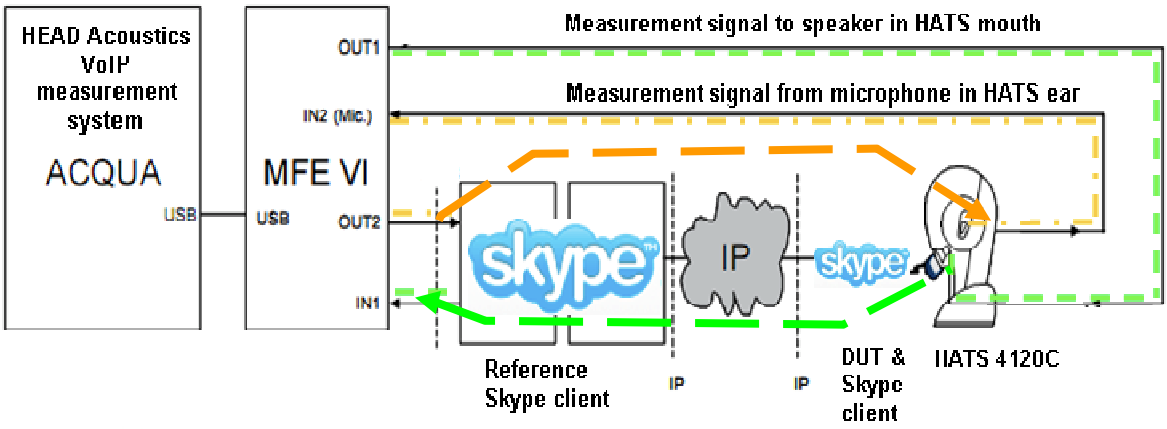
8.1.1 Skype Audio Test Lab

Skype Audio Test Lab consists of state of the art audio testing tools for VoIP and telecommunication:

- Objective testing is performed in a wideband audio measurement capable anechoic room, with a build-in and acoustically „invisible" 18 channel and subwoofer-loudspeaker setup for real and artificial 3D sound reproduction.
- Measurement setup consisting of ACQUA audio testing tool from HEAD Acoustics, with VoIP option and with MFE VI.1 measurement front-end.
- HATS and handset positioner [2] from Brüel & Kjær, HATS model 4128C.
- The actual tests performed by ACQUA system are customized by Skype staff and arranged into test macros, which automate the test process.
- PESQ and other similar advanced objective speech quality models are used in ACQUA system. Skype uses mainly Opticom version of PESQ that has been integrated to ACQUA by HeadAcoustics.
- Free and pressure field measurement microphones and cables from G.R.A.S. Sound and Vibration
- Reference Skype Client on PC with high quality sound card and customizable Skype version
- Several other tools, such as reference headphones: Sennheiser HD650 and HD25, reference low- noise microphone Rode NT2, EMM-8 cal Calibrated measurement microphone, reference loudspeakers: Genelec 8020A, microphone preamps, loudspeaker processor, headphone amp, PC with professional audio editing softwares: Adobe Audition and Audicity and sound cards.
- Skype Certified products will be used as the reference products in the tests.

Copyright © 2009 Skype Inc. All Rights Reserved.

Measurement Setup



Sending direction
Receiving direction
Sending direction delay
Receiving direction delay

- **Head Acoustics ACQUA software – automated audio testing system**
  http://www.head-acoustics.de/eng/telecom_acqua.htm
  http://www.head-acoustics.de/downloads/eng/acqua/acqua18e_mail.pdf
- **Head Acoustics MFE VI.I Measurement Front End**
  http://www.head-acoustics.de/eng/telecom_acqua_mfe_VI_1.htm
  http://www.head-acoustics.de/downloads/eng/mfe/D6462e1_MFE_VI_1.pdf
- **Opticom PESQ (Perceptual Evaluation Voice of Speech Quality) software**
  http://www.opticom.de/download/SpecSheet_PESQ_05-11-14.pdf
- **Bruel and Kjaer Head and Torso Simulator – model 4128C**
  http://www.bksv.com/1650.asp
  http://www.bksv.com/pdf/Bp0521.pdf
- **Bruel and Kjaer Head and Torso Simulator – handset positioner 4606**
  http://www.bksv.com/pdf/Bp0521.pdf
- **Soundcard in Reference Skype Client PC – ECHO Audio MIAMIDI**
  http://www.echoaudio.com/Products/PCI/MiaMIDI/specs.php
- **DUT Skype Client and Reference Skype Client PC specification**
  Intel DG965SS motherboard with BIOS version MQ96510J.86A.1666.2007.0327.2349
  The processor in all PC-s: Intel 630 P4 FSB800 2MB 3.0GHz
  1GB (2x512Mb 533MHz DDR2 NON-ECC CL4 Kingston DIMM modules)
  Samsung 40Gb SATAII NCQ 7200 RPM 8Mb Hard drive
  Samsung DVD ROM
- **Etherfast router Linksys BEFSR81 ver 3.1**
- **WiFi access point for wireless devices – Cisco Aironet AIR-AP1131AG**

Copyright © 2009 Skype Inc. All Rights Reserved.

8.1.2 Compatible testing environment

Manufacturer or audio laboratory willing to test products in their premises need to fulfill at least following conditions and tools:

- ITU-T compatible HATS with type 3.3 anatomic ear, for example from Bruel and Kjaer or Head Acoustics
- Calibrated acoustic and electric measurement system: microphones, amplifiers, wiring…
- Skype recommends to perform measurements mainly at anechoic room as defined in ITU-T documents, for example at ITU-T P.341 recommendation requirement for Test rooms (A.3.1.1), that defines anechoic conditions, sizes of room and noise floor to be below 24 dBA SPL rms
    - It can be possible to use quiet and non-echoic environment/room for headset and handset measurements if the setup is built with care and acoustic knowledge and measurements are performed professionally. In such case Skype considers that noise floor must be below 30 dB SPL A especially for echo and noise floor measurements.
- Care must be put in design and usage of the facility to temporal and stationary noises from: cars, doors, talking, walking, water and drain pipes, ventilation, electricity and radio frequency interferences.

Copyright © 2009 Skype Inc. All Rights Reserved.

9. Appendix

9.1 Definitions

A-weighting	A frequency weighting curve defined in IEC179 and various other standards for use widely in sound level meters. A-weighting is an inverse curve for an equal loudness contour of human hearing at quiet levels (precisely based on the 40-phon Fletcher-Munson curves). A-weighted measurements estimate how people would perceive loudness of a sound taking into account that hearing has different sensitivities for different frequencies.
Acoustic echo	Signal leaking from earpiece or loudspeaker to microphone. For a good call quality this should be as small as possible. If it is too strong it makes communication difficult.
Acoustic user interface	Allows user to hear or speak over the communication system. Products providing acoustic UI have microphone, earpiece or/and loudspeaker. Check Section 1.2.4.
ACQUA	Advanced Communication Quality Analysis system from HEAD Acoustics [11]
Anechoic Chamber / Room	Anechoic chambers are commonly used in acoustics to perform experiments in nominally free field conditions. This means that all sound energy will be traveling away from the source with almost none being reflected back. Anechoic chamber is a room in which there are no echoes.
Annoying sound	A sound that is so clearly audible that it distracts user's attention from the conversation. It can be an unintended consequence of the intended sound or just unwanted sound that irritates the user.
Audible	Means that user can hear certain sound both in quiet environment and in presence of other sounds. Check also *slightly audible*.
Audio ergonomics	Defines how comfortable and meaningful a product is for user from audio perspective, check Section 1.3.
Audio performance	Audible, perceivable performance of a product as judged by the user, check Section 1.3. Consists on sub-parameters – intelligibility, naturalness and conversational effort.
Certification	An endorsement from Skype that a third-party vendor meets Skype's own criteria for a co-labeled solution.
Conversational effort	How little or how much concentration a conversation requires from the user. It is a sub-parameter of audio performance. Check technical parameters from Conversational quality.
Conversational quality	Defines how good or how bad are perceived audio parameters that affect conversational quality. Typical parameters are: delay, acoustic echo, noise and continuity of transmission.
Cordless handset	Handset that operates through radio frequencies without wired connection to a PC or other device. Examples are Bluetooth and DECT handset.

Copyright © 2009 Skype Inc. All Rights Reserved.

Cordless headset	Headset that operates through radio frequencies without wired connection to a PC or other device, for example Bluetooth headset.
Crosstalk	Undesired leaking of receiving i.e. earpiece/loudspeaker signal into sending i.e. microphone signal at transmission circuits.
dB / Decibel	Decibel is a logarithmic representation of a number or ratio between numbers. dB SPL is Sound Pressure Level in decibels (check SPL), dBV is a voltage in dB, so that 1 V (root mean square) equals to 0 dBV.
dBFS	dBFS is a commonly used measure of signal level in digital system compared to decibels full scale. Two different definitions exist, where in both the highest peak level of digital system is 0 dBFS, but RMS level definition varies. The Audio Engineering Society has defined the highest scaled sine signal to have RMS level of 0 dBFS, whereas the other definition sets the same signal RMS level to be -3.1 dBFS. The latter definition is used in some software audio editors. Due to the existence of the two definitions the dBov definition is used for digital levels in this document.
dBm0	Abbreviation for the power in dBm measured at a zero transmission level point. In practice the conversion from dBov is as Y dBm0 ≈ X dBov + 6 dB.
dBov	Measure of a signal level compared to overload point of digital system. Defined in [1] in section 5.7. For a maximum scale digital sine signal the peak level is 0dBov and RMS is -3.1 dBov. dBov definition here is the same as the square wave scaled dBFS definition.
Delay	Delay of speech signal(s) between users
Diffuse field correction	Defined at ITU-T recommendation P.58 [2]. This is the preferred frequency correction of Skype for earpiece and loudspeaker measurements when using HATS. It is a difference, in dB, between the third-octave spectrum level of the acoustic pressure at the ear-Drum Reference Point (DRP) and the third-octave spectrum level of the acoustic pressure at the HATS Reference Point (HRP) in a diffuse sound field with the HATS absent.
Double talk	Situation where two or more parties of a call are talking at the same time.
Electric to electric Skype call	Skype call between two Skype clients that are measured from electric outputs and inputs of good quality sound cards. Acoustic interfaces are not present.
Far end	The other side of the call compared to a local user using device under test. Opposite for the near end.
Free field conditions	Audio measurement environment, with no reflecting surfaces. Such conditions are reachable in an anechoic audio measurement room.
Good quality reference device	A device in every audio UI group that has shown or proven to have good audio quality in various aspects. This device serves as a reference for Device Under Test in some test cases, such as MOS evaluations.
Handset	Product that the user keeps in his hand and puts next to his ear, when in a call, like mobile phone or landline phone, check Section 1.2.2.
HATS	Head and Torso Simulator. Skype Audio Lab HATS is B&K 4128C
Head and torso simulator	Measurement device modeling the head, ear, mouth and upper part of the torso of an average user. Defined in [2].
Headset	Product consisting of earpiece(s) and microphone that the user puts on

Copyright © 2009 Skype Inc. All Rights Reserved.

	his head or ear(s) during a Skype call. Check Section 1.2.2.
Intelligibility	Ability to recognize words and their meanings and also transmission of non verbal information such as emotions, emphasis, identity of speaker. It is a sub-parameter of audio performance.
ITU-T	International Telecommunication Union – Telecom sector, http://www.itu.int/ITU-T/ main standardization organization of speech transmission and quality.
Listening quality	Defines what the perceived quality of speech is. It covers naturalness and partly intelligibility parts of audio performance.
Local user	The person who is using a product under test.
Loudness	Defines a loudness of sound perceived by a listener, other non-scientific everyday life terms are volume, volume level or speech level.
Loudspeaker	Product that converts electric audio signal to acoustic signal – plays back a speech to the user.
Loud speech level	In noisy environment and in situation when people do not hear properly the other participants, people talk at louder level. Technically the level can be even 10 dB A-weighted louder than the normal speaking level. In this document such test signal is used, though the low frequencies have not been amplified with the full 10 dB and crest factor is limited compared to normal speech level.
Lowered speech level	People speak at lowered speech level, when they do not want to disturb other people in the same room. In technical terms the level is around 10 dB SPL lower than the normal speech level, thus the average level in most cases is around 52 dB SPL when measured from 1 m in front of the listener.
Mean Opinion Score	Check definition for *MOS*
MOS	Mean Opinion Score, a numerical indication of the perceived quality of speech or audio. Typically an average of several listeners who have performed a specific MOS test in controlled and formal way. MOS scale is defined in ITU-T recommendation P.800 [3] for speech quality as:

MOS	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

In standardized tests, a good quality narrowband call, for example between mobile phones, can reach MOS slightly above 4. Wideband call can reach close to 5 in good conditions. MOS below 3 is generally considered to be too low.

MOS-LQO	Mean Opinion Score – Listening Quality measured with Objective tools (such as PESQ measurement)
Narrowband speech	Typical landline or mobile phone speech, with a frequency bandwidth between 300 and 3400 Hz.
Naturalness	How natural is listening (and speaking) in conversation. Technical parameters are: adequate loudness, natural frequency content, low noise and distortion. This is a sub-parameter of audio performance.

Copyright © 2009 Skype Inc. All Rights Reserved.

Near end	The side of the call where the device under test is used. Opposite for the far end.
Non-audio product group	Group of products that do not directly influence Skype audio quality during a call, check Section 1.2.5.
Non-acoustic user interface	Products that do not provide acoustic interface.
PESQ	Perceptual Evaluation of Speech Quality tool [10], complies with ITU-T P.862 recommendation. PESQ is used to identify following artifacts: distortion, additional coding, temporal artifacts, additional noise and delay changes, but not acoustic interface artifacts. Skype wants to point out clearly that Skype acknowledges the fact that PESQ has not been designed and verified for acoustic interfaces therefore PESQ is not used as a measure of a quality of acoustic interface. Further Skype uses PESQ as a relative metric comparing the result of an acoustic interface device to a known reference device. In other words Skype is not using PESQ as an absolute metric in acoustic interface cases.
Preferred listening level	Preferred listening levels are defined to be: • 75 dB SPL A-weighted for handset and headset that reproduces speech only to one ear • 69 dB SPL A-weighted for a headset that plays speech to both ears and • 60 dB SPL A-weighted for speakerphone All levels are measured with artificial ear of Head And Torso Simulator [2] when diffuse field frequency correction is applied. The levels are set here based on calculations from ITU-T recommendations for Sending and Receiving Loudness Ratings and other available listening level data. Note that in real life the preferred listening levels between persons can vary up to +/- 10 dB.
Normal speech level	The level/volume/loudness of speech in normal communication between people, in technical terms it is around 62 dB SPL A-weighted when measured 1 m from the user's mouth. At 25 mm in front of the mouth, in so called Mouth Reference Point, the level is defined to be -4.7 dBPa that is 89.3 dB SPL in ITU-T recommendations [4]. In real life this level can vary easily +/-5 dB depending on the person.
Objective testing	Measures quality by means of technical measurement tools.
One-way delay	Delay of acoustic signal from the user to the other party, expressed in milliseconds. If it is below 100 ms then it considered to be good. 400-500 ms delay makes normal conversation difficult.
Other audio product group	Group of products that allow transmission of audio from one system to another; also products that process audio signal, but do not provide acoustic interface to the user, check Section 1.2.4. Examples are: electric audio switch, sound card, Bluetooth dongle.
Other party	The user in a call with a local user (typically not physically located in the same place)
Priority 1	**(Must) requirements:** Priority 1 level requirements are the absolute minimum requirements that the product must pass. For Skype Certification Audio Specification 100% of Priority 1 requirements must pass.
Priority 2	**(Should) requirements:** Priority 2 level requirements are more

Copyright © 2009 Skype Inc. All Rights Reserved.

	demanding. For Skype Certification Audio Specification no less than 50% of Priority 2 requirements must pass.
Priority 3	**(Nice-to-have) requirements:** Priority 3 level requirements show what quality level we would like a Skype Certified product to have. For Skype Certification Audio Specification no less than 10% of Priority 3 requirements must pass.
Product	Part of a solution provided by submitting vendor, includes for example Handset, Cradle, Base station, dongle, application, and drivers, as opposed to a full solution which also includes the latest Skype version.
Purpose	Statement of the requirement that the test case supports.
Receiving	Shortening for Receiving direction or Receiving side, meaning the audio signal path coming from the network to the product and played through the earpiece or loudspeaker to the end user. Simply receiving speech from the other party and playing it to the product user.
Recommended usage position (speakerphone)	For hand-held devices, (such as small phones) the position is in front of the user -30 degree below the mouth at a recommended usage distance, specified by the manufacturer. For non-handheld i.e. desktop devices, recommended usage position is in the middle of a table defined in ITU-T recommendation P.340 [5].
Reference device	Check Good quality reference device
Reference soundcard	Either Echo MIAMIDI or Edirol UA-25 soundcards are used.
RMS	Root Mean Square – a calculation method for average power of signal http://en.wikipedia.org/wiki/Root_mean_square
Round trip delay	Overall acoustic delay of signal from user to the other party and back, synonym to a two-way delay.
Sending	Shortening for Sending direction or Sending side, meaning the audio signal path from the mouth of the user to a microphone of product under test and then transmitted to the other party. Simply sending user's voice to the other party with the product.
Slightly audible	Means that user can barely hear certain sound in a quiet environment. If the user doesn't put any effort he/she might not even notice the sound.
Solution	The product + the latest Skype version
Speakerphone	Product with loudspeaker and microphone, which is usually placed on the table, next to the user during a call. Check Section 1.2.3.
SPL	Sound Pressure Level, expressed in decibels (dB). 0 dB SPL equals to a hearing threshold of silence at 1 kHz tone. SPL is defined as: $L_P = 20 \log_{10}(p/p_0)$, where p is a sound pressure and p_0 is the reference level of 20 µPa. In this document if not else is mentioned SPL refers to RMS power of the signal. SPL of normal conversation varies between 50-75 dB SPL.
Subjective testing	Quality rating based on judgments of test subjects. This requires people to talk or/and listen and rate the quality.
Super wideband speech	Speech transmission, where the audible frequency range is wider than what it is in wideband speech transmission. The sampling frequency is equal or higher than 24 kHz. The bandwidth of signal is between about 50 and 11000 Hz.
THD	Total Harmonic Distortion, a measure of distortion of an audio product.

Copyright © 2009 Skype Inc. All Rights Reserved.

THD+N	Total Harmonic Distortion plus Noise, a measure of distortion of an audio product. Value is typically expressed as percentage, where value 1% (or -40 dB) is known as a limit for inaudibility of distortion of a loudspeaker.
Two-way delay	Overall acoustic delay of signal from user to the other party and back, synonym to round trip delay.
User of the product	User of the product under study.
Vendor	Manufacturer who is submitting a solution for Skype Certification.
Wideband speech	Speech transmission, used in majority of Skype calls, where the audible frequency range is wider than what it is in narrowband speech transmission in PSTN and mobile calls. The sampling frequency is equal to 16 kHz. The bandwidth of signal is between about 50 and 7500 Hz.

9.2 References

[1] ITU-T Recommendation G.100.1: The use of the decibel and of relative levels in speech band telecommunications http://www.itu.int/rec/T-REC-G.100.1/en

[2] ITU-T Recommendation P.58: Head And Torso Simulator (HATS) http://www.itu.int/rec/T-REC-P.58/en

[3] ITU-T Recommendation P.800: Methods for subjective determination of transmission quality http://www.itu.int/rec/T-REC-P.800/en

[4] ITU-T Recommendations P-sector. http://www.itu.int/rec/T-REC-P/en

[5] ITU-T Recommendation P.340: Transmission characteristics and speech quality parameters of hands-free terminals http://www.itu.int/rec/T-REC-P.340/en

[6] ITU-T Recommendation P.57: Artificial Ears http://www.itu.int/rec/T-REC-P.57/en

[7] ITU-T Recommendation P.64: Determination of sensitivity frequency characteristics of local telephone systems http://www.itu.int/rec/T-REC-P.64/en

[8] ITU-T Recommendation G.122: Influence of national systems on stability and talker echo in international connections http://www.itu.int/rec/T-REC-G.122/en

[9] ITU-T Recommendation G.131: Talker echo and its control http://www.itu.int/rec/T-REC-G.131/en

[10] Perceptual Evaluation of Speech Quality tool, PESQ that complies with ITU-T P.862 recommendation http://www.opticom.de/download/SpecSheet_PESQ_05-11-14.pdf Skype uses Opticom version of PESQ that has been integrated into HeadAcoustic ACQUA system [[11]]

[11] ACQUA Advanced Communication Quality Analysis system by HeadAcoustics http://www.head-acoustics.de/eng/telecom_acqua.htm

9.3 Changes between 4.0 and 3.0 versions

9.3.1 Major changes

Audio ergonomics requirements are removed from the 4.0 version and similar requirements are added to the other Certification documents. Thus in all requirement chapters of the version 4.0 the Sections “Audio ergonomics requirements” have been removed.

The final pass criteria of device tested against this document has been relaxed for the 4.0 version. In version 4.0 it is that 100% of Priority 1 requirements must be passed, 50% of Priority 2 and 10% of Priority 3. The corresponding percentages in the version 3.0 and earlier versions are 100%, 75% and 25%. The reason for the relaxation is that the old pass criteria required a product to pass a considerable amount of Priority 2 and 3 requirements, and if the product fails these that would mean considerable and time-consuming hardware changes for product in a middle of the development cycle.

Copyright © 2009 Skype Inc. All Rights Reserved.

The specification has started to use Diffuse field correction instead of Free field correction in earpiece and loudspeaker measurements in all acoustic measurements when using artificial ear of HATS. This is changed to reflect the latest information in research regarding the preferred frequency response targets in subjective user tests and the latest developments in ETSI and ITU-T standardization forums. However Skype does not state that flat diffuse field corrected frequency response is the optimal when measured with HATS, the frequency mask allow plenty of room for manufacturers to optimize the response.

Super wideband frequency response requirements have been added aiming for frequency ranges beyond wideband to allow very high quality speech, music and multimedia delivery.

9.3.2 Introduction, Abbreviations and References

Few minor text editions added to “Introduction” chapter. Example pictures for Acoustic UI groups have been updated, for example sound cards added to “Other audio product group” section.

Text in “Audio requirements and priorities – overview” section has been modified to reflect the fact that Audio ergonomic requirements have been removed from this document.

The new final pass criterion is presented in the end of section “Use of the test case priorities”.

New abbreviations are added: ACQUA, Diffuse field correction, Far end, MOS-LQO, Near end, PESQ, RMS, and Super wideband speech. Also few abbreviations have been updated, for example: preferred listening level, Priority 1, 2 and 3 and SPL.

Two more references have been added to “References” section: PESQ and ACQUA.

9.3.3 General audio requirements

Major updates have been made to this Chapter.

“Additional delay to speech signal…” requirements for sending and receiving directions in Version 3.0 have been combined to new “Round trip delay of speech signals” requirements. The “Additional delay…” requirements have been removed. The measurement method has been redefined and requirements updated.

Several requirements in the version 3.0 have been combined into only two test requirements: “Total quality loss in sending direction” and “Total quality loss in receiving direction”. The combined old requirements are:

- Format and additional coding of speech – all priorities here
- No drops and distortions in speech signals – all priorities here
- No additional noises or sounds in speech signals – all priorities here
- No interference noises from electric power supply – all priorities here
- No interference noises from devices with radio frequency transmission – all priorities here

All of these combined requirements are removed from the 4.0 version.

“Frequency bandwidth…” requirements have been removed from the 4.0 version. The bandwidths are tested in Frequency response requirements in the version 4.0.

“Device and driver response time from the audio mixer” requirement has been relaxed from 1 ms in version 3.0 to a practical 50 ms in the version 4.0. Also the text has been updated.

“Sampling frequency accuracy” requirement has been relaxed from 0.01% i.e. 100 ppm deviation in the version 3.0 to 0.1% i.e. 1000 ppm in the version 4.0. Also the text has been clarified.

“General audio test instructions” section has been updated to reflect changes in the test cases. The “subjective way” section has been removed as all tests in the Chapter are measured with objective tools.

Copyright © 2009 Skype Inc. All Rights Reserved.

9.3.4 Headset audio UI

The “Microphone - frequency response” requirements have been relaxed from high frequencies by adding 5dB more headroom to the tolerance window in the version 4.0. For Priority 2 requirement the low end corner frequency of the tolerance window has been raised from 100 Hz to 150 Hz. Priority 3 requirement has been added to test a super wideband compatibility of the microphone.

Priority 2 “Microphone – Speech to background noise ratio” requirement has been rewritten for better clarity. Priority 1 and 3 requirements of the same test have been removed.

All priorities of “Earpiece – Speech to self noise ratio” requirements have been tightened by 5 dB in the version 4.0 compared to the version 3.0.

In “Earpiece – Frequency response” Priority 2 requirement, the low and the high corner frequencies have been both lowered from 300 and 15000 Hz to 100 and 7000 Hz to the version 4.0. Priority 3 requirement tolerance mask is also modified: high frequency has been dropped to 10 kHz in the version 4.0 compared to the previous value of 20 kHz, due inaccuracies of practical measurement with artificial ears. On the other hand the low frequency tolerance mask has been tightened due change to diffuse field correction and more knowledge about the preferred frequency response there.

In “Earpiece – Stability of frequency response” the frequency limits have been slightly changed to reflect the modified tolerance mask frequencies of Frequency response requirements. The modified limits here are: Pr 2: the highest frequency is lowered from 7 to 6 kHz, Pr 3: the highest frequency is lowered from 8 to 7 kHz, and Pr 3: the lowest frequency is raised from 100 to 150 Hz.

“Minimum crosstalk from receiving to sending direction” requirement has been modified to use A-frequency weighted values in the version 4.0 instead of not-weighted values in the previous versions.

“Headset: Audio ergonomic requirements” section has been removed from the version 4.0.

“Headset: Supporting audio documentation requirements” has been updated and only the most important information for Certification testing purposes has been left.

“Headset: Audio test instructions” section has been updated to reflect changes in test cases. The “Subjective way” section has been removed as all tests in the Chapter are measured with objective tools.

9.3.5 Handset audio UI

The “Microphone - frequency response” requirements have been relaxed from high frequencies by adding 5dB more headroom to the tolerance window in the version 4.0. For Priority 2 requirement the low end corner frequency of the tolerance window has been raised from 100 Hz to 150 Hz.

Priority 2 “Microphone – Speech to background noise ratio” requirement has been rewritten for better clarity. Priority 1 and 3 requirements of the same test have been removed.

All priorities of “Earpiece – Speech to self noise ratio” requirements have been tightened by 5 dB in the version 4.0 compared to the version 3.0.

“Earpiece – Frequency response” requirements are moved before “Earpiece – Suitable volume level…” requirements in the version 4.0.

In “Earpiece – Stability of frequency response” the frequency limits have been slightly changed to reflect the modified tolerance mask frequencies of Frequency response requirements. The modified limits here are: Pr 2: the highest frequency is lowered from 7 to 6 kHz, Pr 3: the highest frequency is lowered from 8 to 7 kHz, and Pr 3: the lowest frequency is raised from 100 to 150 Hz.

“Minimum crosstalk from receiving to sending direction” requirement has been modified to use A-frequency weighted values in the version 4.0 instead of not-weighted values in the previous versions.

Priority 3 requirement for “Suitable volume level for office and home handset (Indoor)” has been removed from the version 4.0.

Copyright © 2009 Skype Inc. All Rights Reserved.

"Handset: Audio ergonomic requirements" section has been removed from the version 4.0.

"Handset: Supporting audio documentation requirements" has been updated and only the most important information for Certification testing purposes has been left.

"Handset: Audio test instructions" section has been updated to reflect changes in test cases. The "Subjective way" section has been removed as all tests in the Chapter are measured with objective tools.

9.3.6 Speakerphone audio UI

Microphone and echo requirements are moved to be before the loudspeaker requirements in the version 4.0.

"Priority: 2 Microphone – Frequency response" the low frequency of the tolerance mask has been increased from 100 to 150 Hz for the version 4.0. Priority 3 requirement has been added to test a super wideband compatibility of the microphone.

"Microphone – Speech to self noise ratio during speech activity" requirement has been made stricter by increasing A-weighted RMS speech to noise ratio to be at least 30 dB in the version 4.0 compared to 25 dB in the version 3.0.

Texts in "Amount of acoustic echo" requirements have been clarified and Notes have been updated.

In "Loudspeaker – Frequency response" requirements for frequencies in tolerance limits have been modified: Priority 1: the high frequency limit is reduced from 3.5 to 3.4 kHz, Priority 2: the high frequency limit is reduced from 7.5 to 7 kHz, and Priority 3, the high frequency limit is reduced from 15 to 10 kHz.

For "Loudspeaker – Distortion…" requirements, the text has been clarified and example is given how to define the measurement bandwidth for a distortion measurement.

"Microphone – Sensitivity at maximum operating distance" requirement has been clarified.

"Sampling frequency accuracy – absolute" requirement in the version 3.0 has been removed from the version 4.0.

"Speakerphone: Audio ergonomic requirements" section has been removed from the version 4.0.

"Speakerphone: Supporting audio documentation requirements" has been updated and only the most important information for Certification testing purposes has been left.

"Speakerphone: Audio test instructions" section has been updated to reflect changes in test cases. A graph defining the measurement setup has been added.

9.3.7 Other audio product

"Product provides suitable levels for audio signal output" requirement has been updated and reference acoustic UI products have been added.

A new requirement has been added: "Product provides suitable levels for audio signal input". It also includes reference acoustic UI products.

"Minimum crosstalk from receiving to sending direction" requirement has been modified to use A-frequency weighted values in the version 4.0 instead of not-weighted values in the previous versions.

"Other audio product: Audio ergonomic requirements" section has been removed from the version 4.0.

"Other audio product: Supporting audio documentation requirements" has been updated and only the most important information for Certification testing purposes has been left.

"Other audio product: Audio test instructions" section has been updated to reflect changes in test cases. The "Subjective way" section has been removed as all tests in the Chapter are measured with objective tools.

Copyright © 2009 Skype Inc. All Rights Reserved.

9.3.8 Non-audio product

Texts in both requirements of “Continuous transmission of speech” have been clarified.

The “Subjective way” section has been removed as all tests in the Chapter are measured with objective tools.

9.3.9 List of environments

“List of test tools and material” section has been removed due less need to use subjective measurement and listening setups in grading requirements in the version 4.0 of Certification audio requirements.

“Skype Audio Test Lab” has been updated. More details of equipment used in testing are defined also a graph describing the objective measurement setup.

Copyright © 2009 Skype Inc. All Rights Reserved.